MIDDLESEX UNIVERSITY

Molecular Assessment of Parasite Infection within Socioeconomically Important UK Salmonid Populations

A thesis submitted to Middlesex University in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Toby Landeryou

M00513028

Faculty of Science and Technology, Department of Natural Science
Hendon Campus, London, England

Acknowledgements

I would like to first and foremost thank my supervisory team of Dr Steve Kett, Dr Dirk Wildeboer and Dr Scott Lawton with helping me out throughout the planning, practical and most importantly writing of my thesis. Also although not part of the official supervisory team, Dr Anne Ropiquet has been a constant guiding hand, taking time out of her own work to aid me throughout the project and for that I am extremely grateful. They have all gone above and beyond in helping me achieve the final completed copy of this work and I am extremely thankful for their due diligence and great attention to detail when helping me through to completion.

Alongside the academic supervisory there has been a substantial input to create the extensive archive at Middlesex University, which was vital to this project. For this I would the Wester Ross Fishery Trust, a truly great bunch of people to work with who are dedicated to maintaining a sustainable Gairloch system that can be enjoyed by anglers and naturalists for generations to come. The contributors have been many but to mentions a few I would like to thank the generous list of highly skilled and adept anglers who made the archive possible; Dr Steve Kett, Dr Andy Vick, Vu Dang, Dr James Close, Dr Robin Ade, Prof Barry Blake, Ben Rushbrooke, Prof Dave Barclay, David, Dougal and Florence Foremen, Frank Buckley, Hugh Whittle, Prof Pete Maguire, Pat Wilson, , Dr Carl Forbes, Ray Dingwall, Ronnie Buchan Roger McLachlan, Robin Payne, Radenko Velinov, Calum Button and Adam Choonara.

Most importantly I would like to thank my caring and patient wife for putting up with constant late nights and misguided thesis-based tantrums. Also my mother who has always been there to keep me on the straight and narrow and to fill my stomach full of home baked goods.

A man who this thesis is dedicated to is my beloved father who passed away on the 16th of August 2018 after a brief but courageous fight against cancer. I will always be indebted to a man that always put the needs of his family before the needs of his own. Like his son he was an average but enthusiastic fly fisherman, whose snoozes in the sunshine will vastly outweigh the time spent with fish on line. Although you won't see me walk on stage in a "poncy hat" you will be in my mind every step and every day of my journey in the world science and life. Love you Dad and thank you, your son did good.

R.I.P Bryan Landeryou

1957-2018

Abstract

Salmo trutta (Brown trout) is a native salmonid species to the United Kingdom. The economy generated via fish farming and eco-tourism has aided isolated communities in job creation and tourism income. It is particular true in highland communities in Scotland, where parasite infection has threatened the health of Scottish salmonids. The work presented in this thesis assessed parasite infection of wild brown trout populations in the northwest highlands of Scotland. Initial analysis of fish health screened subpopulations within the Gairloch region for infective parasite species. Eye fluke *Diplostomum* spp. was highly prevalent throughout the system with all lochs populations harboring infection. Through molecular analysis the species was confirmed as *D. baeri* which, when compared to other European isolates indicated highly diverse species complex infecting freshwater fish across the continent, most likely through definitive bird host. A cestode species that infected 4 trout subpopulations across the system was the medically relevant *Diphyllobothrium dendriticum*. Using molecular species identification techniques, it was the first finding of the parasite within UK freshwater fish populations. In response to differential parasite infection adaptive immunogenic traits were also observed within trout host subpopulations. The MHC II related gene *Satr*-DAB variability was higher in populations with differential parasite species infection suggesting the diversity of infection maintains MHC diversity within the population. The evasion of immune recognition to achieve high levels of infection is key to sustained parasite infection. The highly infective parasite in the Gairloch system, *D. baeri,* utilizes intracellular antigens Tetraspanins and Venom allergen-like antigenic proteins. These antigens were isolated using genomic techniques to highlight potential vaccine targets in aquaculture. The body of work presented here has furthered the knowledge of the highly infective *D. baeri* and provided molecular methodologies to identify medically relevant *D. dendriticum*. Genomic analysis of trout population immunogenics and parasite antigenic factors provides key knowledge to further conservation stocking methods and sustainable aquaculture practice.

Contents

Table of Figures

---

1 Introduction

1.1 Brown trout ecology

Brown trout, *Salmo trutta* L. 1758 is a salmonid fish indigenous to Europe, North Africa and western Asia. Within Europe its northern limits reach as far as Iceland, the northern territories of Scandinavia and Russia (MacCrimmon *et al.*, 1970). Introduction has occurred in numerous countries outside the trout's native range, with successful introductions occurring within 24 countries outside of its indigenous environment (Louis Bernatchez, 2001). Global success stems from its ability to spread and colonize a variety of different habitats from small highland streams to large arctic fjords. Adaptation to various environments has led to similarly various phenotypic appearance, with more than 50 different variations of the brown trout being described in early speciation studies and with all species of brown trout currently grouped together in a large polymorphic species complex (Louis Bernatchez, 2001). Within one population of brown trout, members may occupy running water spawning habitats, large lake residential populations, brackish water and, in its anadromous form (sea trout, *S. trutta trutta*), local coastal waters (Neil, 2013). Few other European fish have this ability to inhabit such a wide variety of freshwater habitats, largely enabled by brown trout's migratory behavior and wide range of environmental tolerance (Klemetsen *et al.*, 2003). Riverine trout feed on insect larvae, highly dependent on insect hatch time of year and geography (Gíslason and Steingrímsson, 2004). As they grow in size their choice of insect prey changes concomitantly; juveniles favoring small, surface-dwelling arthropods, larger trout hunting insect larvae from larger insect groups such as Plecoptera, Trichoptera and Ephemeroptera (Rincon and Lobon-Cervia, 1999). Lacustrine trout populations also exhibit age-variable diet, juveniles feeding on shallow water zoo benthos, zooplankton and insects of varying size dependent on the size of fish (Keeley and Grant, 2001). Brown trout populations are often known to be piscivorous, particularly the phenotypic variant of brown trout known as 'ferox' (Jardine, 1835) (L'Abee-Lund *et al.*, 1992). Generally, however, trout exhibit highly opportunistic feeding habits depending on numerous environmental factors including age and size. Brown trout populations inhabit

complex systems with variability in life history and feeding habits key to their adaptation to a myriad of different habitats with self-sustaining populations existing worldwide.

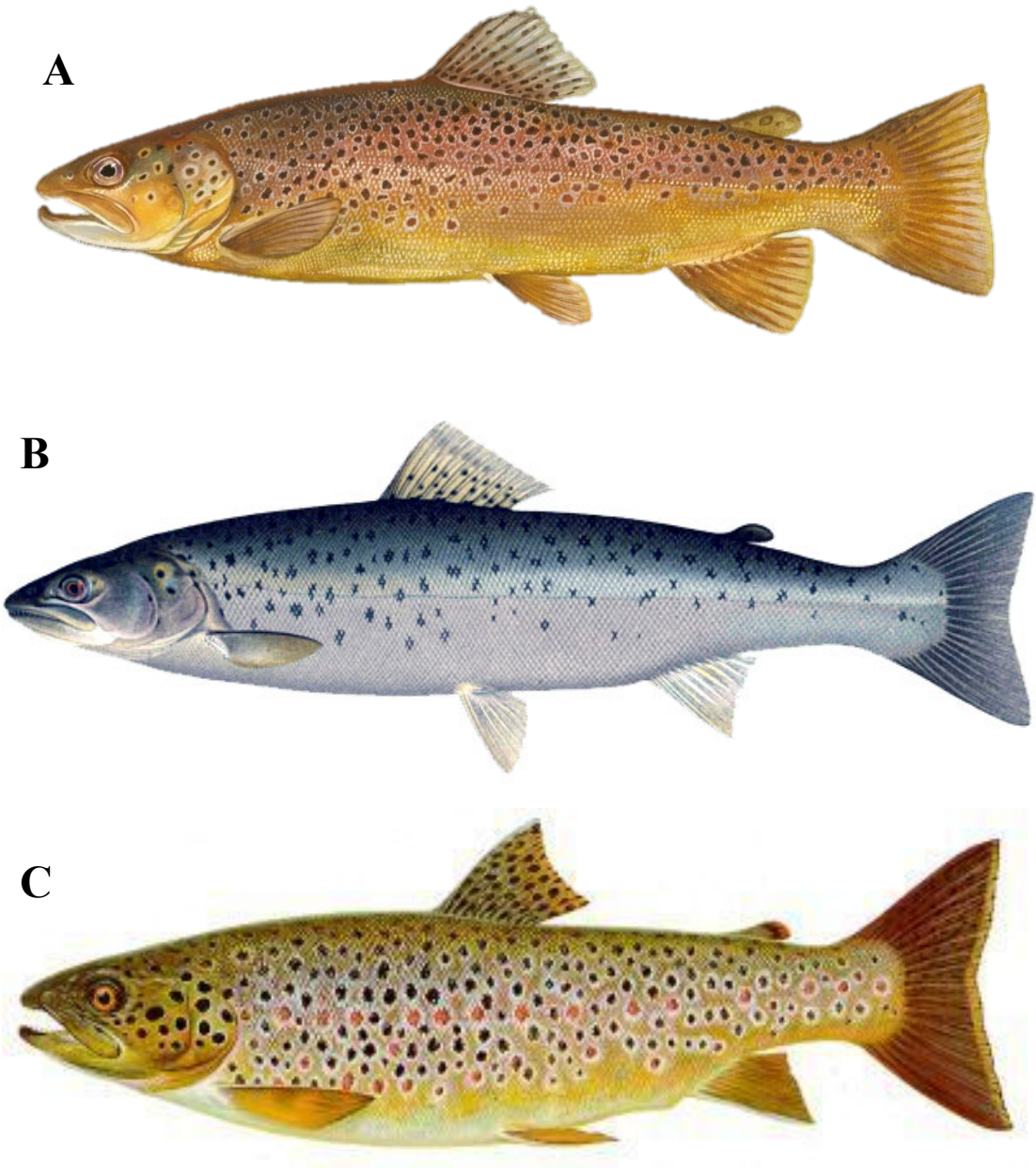


Figure 1: Detailed diagram of *S.trutta* morphs existing across Gairloch freshwater system. **A** *Salmo trutta fario,* lacustrine brown trout. **B**, *Salmo trutta trutta;* anadromous sea trout morph of brown trout. **C** *Salmo ferox,* carnivorous morph of brown trout more commonly seen in deep loch populations of brown trout (Raver, US Fish and wildlife, 2012)

1.1.2 Brown trout health

Within the environments brown trout inhabit they are linked to many food webs, acting as food source for predatory birds and mammals (Neil, 2013). Being both a popular food source and highly successful in many climates, brown trout is also an integral part of many parasite species life cycles (Criscione and Blouin, 2004; Klemetsen *et al.,* 2003). The effect that parasite infection has on host populations of brown trout can vary between species of parasite. The cestode genus *Eubothrium* is widespread in trout populations but causes little detriment to overall fish health (Boyce and Yamada, 1977). In contrast the monogean parasite genus, *Gyrodactilus* has caused widespread trout death across Scandinavia (Buchmann and Uldal, 1997). In recent years identification of trout parasites has garnered attention due to select species' endangered status including Gila trout, *Oncorhynchus gilae* (R.R. Miller 1950) (Peters and Turner, 2008), Yellowstone cutthroat, *Oncorhynchus clarkia bouvieri* (Jordan and Gilbert, 1883) (Ruzycki *et al.*, 2003) and Apache trout, *Oncorhynchus apache* (R.R. Miller, 1972) (Petre and Bonar, 2017). Brown trout is a species of least concern to conservation worldwide, however its use in aquaculture and sports angling make the health and stocking practice of socioeconomically important populations vital to communities relying on such economies (Whitmarsh and Wattage, 2006).

1.1.3 Brown trout-parasite relationship

Within freshwater environments there are a multitude of potential host species making it ideal for parasites to complete full life cycles within a very short geographical area (Medel *et al.*, 2010). Taxon present in many freshwater systems are freshwater gastropods, with genera such as *Bulinus, Lymnaea* and *Planorbis* acting as crucial primary hosts in a variety of freshwater digenea parasites (Gíslason and Steingrímsson, 2004). Other primary host components of parasite life cycles can include oligachete worms and copepods (Gíslason and Steingrímsson, 2004). Brown trout are opportunist feeders, with infection frequently occurring through the direct consumption of copepod and stickleback primary hosts involved in cestode life cycles. Freshwater trematodes instead complete life cycle by utilizing a multitude of freshwater hosts instead of direct consumption of parasite by the host. Freshwater trematode lifecycles frequently involve

an infected snail hosts releasing a free-living parasitic cercariae that attaches and penetrates trout musculature before reaching immunogenically inert infection site. Parasite infective potential varies throughout brown trout life cycle, in both temporal (Tully *et al.*, 1999) and spatial (Vehanen *et al.*, 1999) facets. The wide variety of environments inhabited during the lifetime of a brown trout can drastically vary the infective potential of parasites. Trout populations inhabiting slower moving water bodies experience increased cercarial presence (Buchmann and Bresciani, 1997), whilst populations inhabiting larger, deeper lakes will consume larger prey such as oligachetes, are thus frequent hosts of nematodes such as *Eustronglyides* spp. (Hirshfield *et al.,* 1983*)*. Pathogenicity varies between parasite species, however, virulent infection of trout fry can lead to wide scale fish death and population decline (Rahkonen *et al*., 1996). Cestode infection can vary in pathogenicity; *Eubothrium* spp. are widespread cestode parasites of trout, with fish exhibiting even large infection numbers showing little ill effect (Hanzelová *et al.*, 2002). Cestode infection are not always benign in nature with large-scale epidemics of *Diphyllobothrium* spp. being responsible for killing large numbers of trout fry in reservoirs across England (Fraser, 1960). Nematode infections within trout exhibit a similar lifecycle to those of cestode species with trout infected via direct consumption of infected intermediate hosts such as oligachete worms (*Eustronglyides* spp.) and mayfly nymphs (*Sterliadochona* spp.) (Frimeth, 1987). Unlike cestodes, nematode infections in trout are not limited to the intestinal tract, with taxa like *Contracaecum* spp. being found within the atria of rainbow trout in numerous habitats with high pathogenic potential (Oscoz *et al.*, 2005.). Due to migratory movements of trout, the fish can also become infected by marine nematodes such as *Anasakis* spp. that usually infect sea trout and estuarine populations of riverine brown trout (Wootten and Smith, 1975). Freshwater trematodes present a different interaction with trout host species with infection frequently involving a behavioral modification (Barber *et al.*, 2000). *Diplostomum* spp. are eye flukes that utilize brown trout as an intermediate host, residing within the trout's ocular lens during part of its lifecycle. Pathogenicity in heavy infections involves partially blinding the fish and increasing probability of predation (Grobbelaar *et al.*, 2015). *Tylodelphys* spp. infects the brains of freshwater fish, affecting their feeding habits and movement (Moody and Gaten, 1985). Other intriguing parasite

species that infects freshwater and anadromous trout are the haemoflagellate species; *Cryptobias* and *Trypanosoma* species. Both haemoflagellates complete their early life cycles within leeches (Hirudinea) and pass the trypomastigotes stage to the fish host during blood feed (Ardelli and Woo, 2001). The epidemiology and infection dynamics of both parasite species are poorly understood but hydrology plays an important role in leech and fish interaction (Fermino *et al.*, 2015).

As with other aspects within a brown trout's life history, parasitism exerts influence, not only at the individual level, but also at the population level. Many of the differential parasite infections occurring between populations of trout are related to host movement (Ziętara *et al.*, 2007). The ecological impact of host transplacement can drastically affect overall host population health. Invasive trout species can succumb to novel parasite infection within new environment adding to the environmental complexity populations of brown trout must overcome to thrive (Monzón-Argüello *et al.*, 2014). Alternatively, populations can undergo release from pathogenic pressure if they are translocated to zones in which parasites/pathogens are absent.

The success of brown trout populations in varied freshwater landscapes makes them an ideal host for a parasite infection. Their ability to migrate, survive off multiple food resources and their role as prey for multiple definitive host species make completion of a parasite life cycle highly probable. Equally, the ability to withstand parasite infection is a key factor in maintaining health within brown trout populations.

1.2 Diplostomum

1.2.1 Life cycle and pathology

The oft-parasitized brown trout is infected by a variety of parasites due in part to its ability to survive within varied environments. A parasite commonly found in a variety of riverine and lacustrine population of brown trout is the eye fluke; *Diplostomum* spp. Trout act as intermediate hosts for the metacercarial stage which resides within the oculus of the trout (Betterton, 1974; Lyholt and Buchmann, 1996; Stables and Chappell, 1986a; Whyte et al., 1987, 1990, 1991). Infection is achieved via penetration by free-living cercariae, shed by infected freshwater snails (Lyholt and Buchmann, 1996). Snail infection occurs through the consumption of sporocysts, shed from definitive hosts into the water. Once the sporocyst is consumed, the parasite migrate to the gonads of the snail, making the host immunogenically inert, ensuring increased duration of infection and prolonged cercarial production (Karvonen *et al.*, 2006). Released cercariae exhibit a free-living life span of approximately 8 hours during which they must attach to the mucosal layer of the trout host, with motility directly associated with water temperature (Karvonen *et al.*, 2003). Post cercarial penetration *Diplostomum* migrate through host musculature to infection site within the immuno-inert ocular site (Bakal *et al.*, 2005) (figure 2). Pathology associated with intermediate host infection is recognizable cataract formation (Larsen *et al.*, 2005). Within wild fish populations infection accumulates annually due to high cercarial prevalence months, coinciding with warmer water temperatures in summer and early autumn (Stables & Chappell 1986; Larsen *et al.* 2005; Turgut and Ozgul, 2012). In severe cases, heavily infected fish suffer from blindness, emaciation and death. Histopathology associated with severe ocular damage can include exophthalmia, lens dislocation, capsular rupture and retinal detachment. Mortalities resulting directly from infection are difficult to ascertain due to death being associated with secondary effects from the infection (Kristmundsson and Richter 2009; Larsen *et al.* 2005).

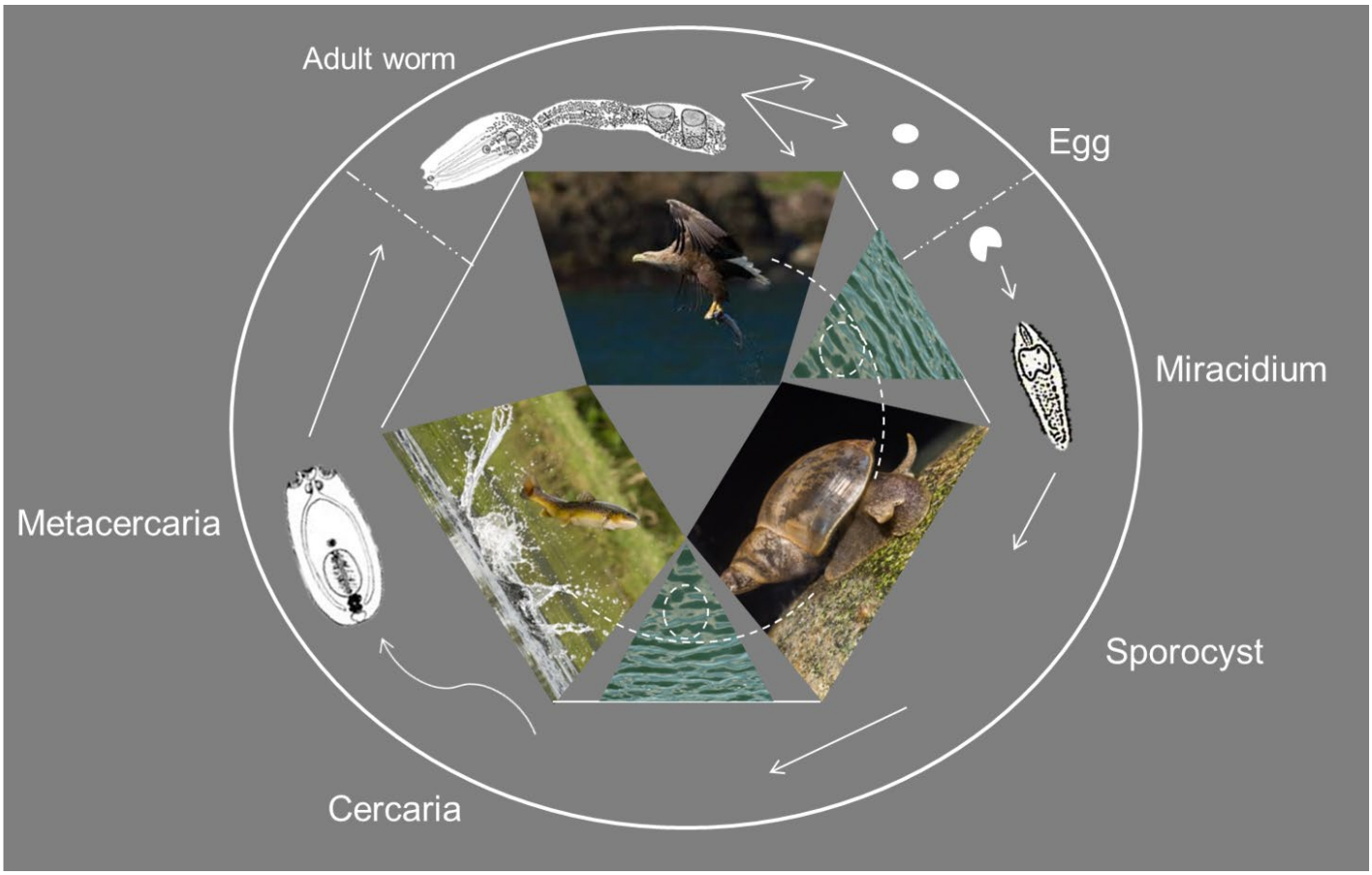


Figure 2; Life cycle of *Diplostomum* spp.

1.2.2 Taxonomy of *Diplostomum*

Eye flukes are some of the most ubiquitous infective parasite taxa. *Diplostomum spathaceum* has been found infecting over 100 species of fish worldwide (Chappell, 1994). The geographic infection range of *Diplostomum* is wide, found infecting hosts within Canada (Brassard *et al.*, 1982; Marcogleise *et al.*, 2001; Galazzo *et al.*, 2002), Europe (Buchmann, 1985; Ballabeni, 1994; Pylkko *et al.*, 2006; Georgieva *et al.*, 2013), East coast of Africa (Migiro *et al.*, 2012; Ndeda *et al.*, 2013) and the Sub-Arctic (Frandsen and Malmquist, 1989; Voutilainen and Taskinen, 2009; Blasco Costa *et al.*, 2014).One of the main reasons for widespread prevalence is the range of hosts utilized within *Diplostomum* lifecycle, and high mobility of the definitive host species; migratory piscivourous bird species (Karvonen *et al.*, 2005). With the parasites widespread infective potential investigation into taxonomy and species identification is an important tool to ascertain geographical species diversity.

The taxonomic identification of freshwater parasites has become an increasingly important aspect of parasitology, particularly in assessing the impact of infection in wild populations of host species (Faltýnková *et al.*, 2016; Poulin, 2016; Poulin and Mouillot, 2003). Early identification of *Diplostomum* spp. proved problematic providing mixed results due to their morphological similarity (Niewiadomska, 1996). Identification worldwide has benefitted greatly from incorporation of molecular techniques, which aid taxonomic identification of previously ambiguous species, leading to the discovery of cryptic species in the genus (Vilas *et al.*, 2005;Blasco-Costa *et al.*, 2014; Kudlai *et al.*, 2017).

Molecular information has been acquired for 19 species of *Diplostomum* spp*,* including the previously morphologically identified *Diplostomum baeri* (Dubois, 1937), *Diplostomum huronense* (La Rue, 1937), *Diplostomum spathaceum* (Rudolphi, 1819) and *Diplostomum indistinctum* (Niewiadomska, 1984). Molecular taxonomic studies on European isolates are relatively recent (Georgieva et al., 2013a; Selbach et al., 2015a; Faltýnková et al., 2016; Kudlai et al., 2017a) discovering 12 species across northern (Blasco-Costa *et al*., 2014a), central (Georgieva et al., 2013a; Kudlai et al., 2017a;

Selbach et al., 2015a) and southern (Pérez-del-Olmo *et al*., 2014a) Europe. This included the discovery of a species complex structure within *D. baeri* (Isabel Blasco-Costa *et al*., 2016) and *D. mergi* (Selbach et al., 2015a) isolates. The wide ranging infection of *Diplostomum* is highlighted in the discovery of the species within multiple tropical locations; Australasia (Cribb *et al.*, 1993.), South America (Szidat, 1969) and Africa (Musiba and Gamba, 2006). As indicated by the dates, these studies were performed early in taxonomic identification methodology and based purely on morphological differentiation. With the onset of molecular taxonomic techniques a fuller picture of *Diplostomum* taxa is becoming clear. Worldwide it consistently presents as a highly speciated parasite, with the ability to infect multiple freshwater fish species. With the increase use of molecular identification techniques within varied localities, the true estimation of species diversity and global richness can be achieved.

1.3 Evolutionary ecology of the MHC complex

1.3.1 MHC structure and function

One key mechanism that allows local adaptation of host populations is the adaptation of host immunogenetics (Sommer, 2005). Key genes associated with the host immune system (Zueva *et al.*, 2014); the major histocompatibility complex (MHC) plays a vital role in host infection defense (Piertney and Oliver, 2006). MHC genes encode for proteins that present pathogen-derived antigens to T-cells initiating an adaptive immune response (Simpson, 1988) and are amongst the most investigated functional genes in vertebrates (Hughes and Yeager, 1998). The MHC gene family encompasses two main sub-groups of immunological molecules. Class I molecules are expressed on the surface of all nucleated cells and present intracellular derived peptides to CD8+ cytotoxic T-cells. The Class I MHC cells are primarily associated with defense against viral and bacterial insult (Simpson, 1988) . Class II molecules are present on antigen presentation cells like macrophages and lymphocytes and process exogenous antigens to CD4+ T-helper cells. Class II molecules are associated with the immune defence of parasites and pathogens.

To recognize the foreign pathogen MHC structure includes a “basket” receptor called the “antigen binding site” (ABS) (Cuesta *et al.*, 2006; Cutrera *et al.*, 2014; Forsberg *et al.*,

2007; Medel *et al.*, 2010). The role that the ABS plays in direct interaction with foreign pathogens makes it a target of studies focusing on how parasite-driven selection can facilitate an adaptive immune response at a genetic level in vertebrate populations (Eizaguirre *et al*., 2012; Lamaze *et al*., 2014a; Natsopoulou, 2010). Although ABS regions do show some degree of specificity a single MHC molecule can recognize and bind to multiple pathogen peptides that share particular amino anchor positions (Altuvia and Margalit, 2004). Genetically MHC molecules are some of the most polymorphic genes in vertebrates with the largest variability being exhibited within the ABS region (Schenekar and Weiss, 2017) (Figure 3). Genetic theory suggests that at a population level resistance/tolerance of a diverse array of pathogens is associated with high MHC diversity across a host population (Sommer, 2005). A large body of empirical data indicates selection via contact with pathogens acts on MHC loci to maintain MHC diversity within populations (Radwan *et al.*, 2014; Bracamonte *et al.*, 2015: Perchouskova *et al.*, 2015; Schuster *et al.*, 2016).

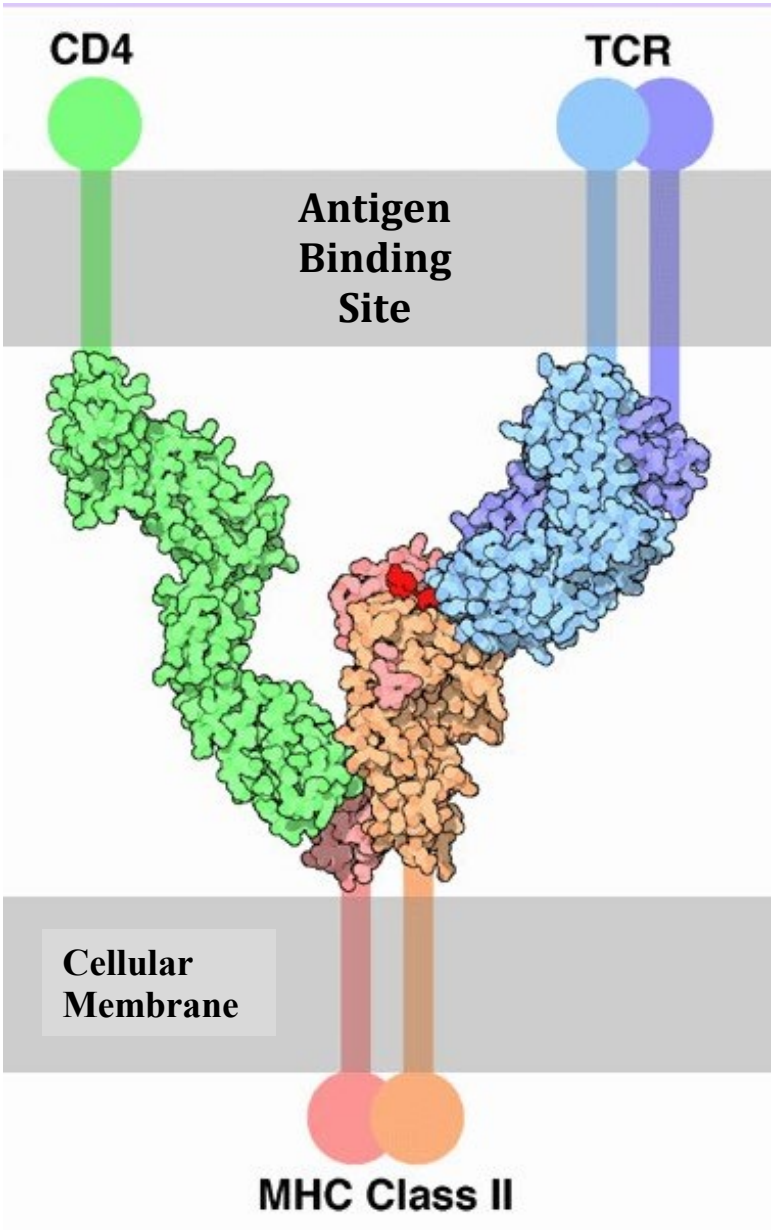


Figure 3; MHC II schematic diagram, Antigen binding site (ABS) is labeled within binding cleft, the site of direct protein pathogen interaction. TCR and CD4 labels denote regions of MHC II that interacts with further CD4+ T-cell receptor to enact further immune cascade (Goodsell, 2013)

1.3.3 Salmonid MHC ecology

With wild populations of salmonids under varied conservation status (Waples and Hendry, 2008) and their increased farmed fish usage, salmonid health is of high socioeconomic importance. Numerous studies have investigated the underlying molecular basis of the salmonid immune response to infection (Consuegra and Garcia de Leaniz, 2008; Zueva *et al.*, 2014). The diversity of salmonid MHC complex is viewed as an important factor contributing to overall population health (Miller and Withler, 2004). For salmonids genetic variability of the MHC loci has been attributed to a number of different factors including; mate choice (Consuegra and Garcia de Leaniz, 2008), embryo and juvenile fitness viability (Forsberg *et al.*, 2007) and pathogen infections (Lamaze *et al*., 2014a) and, thus, also the long term viability of invasive species or non-native strains (O'Farrell et al., 2013a; Schenekar and Weiss, 2017). The negative effects that stocking has on MHC diversity within wild systems has also been well documented (Currens *et al.*, 1997; Lamaze *et al.*, 2017).

Evolution and maintenance of the MHC polymorphism in salmonids in response to pathogen infection has proven a major mechanism in local adaptation and population fitness (Buchmann and Uldal, 1997; D. J. Fraser et al., 2011; Lamaze et al., 2014a). The wide body of MHC-associated research has addressed pathogen-driven selection in various pathogens; anemia virus (Mjaaland *et al.*, 2005), *Aeromonassalmoncida* bacteria (Croisetière *et al.*, 2008), *Gyrodactylussalaris* parasite (Tonteri *et al.*, 2010) and stocked *vs* wild population adaptation (O'Farrell et al., 2013a; C. Monzón-Argüello et al., 2014; Schenekar and Weiss, 2017). However studies of the evolutionary response of wild populations to specific parasite infection are sorely lacking, with the majority of research based primarily on inferred parasite infection rather than parasite screen derived data.

1.3.2 MHC ecology and parasite driven selection

The determination of the relative role that microevolutionary processes play in the maintenance of genetic diversity within natural populations has been a major focus of pathogen-host interaction studies for years (Eizaguirre *et al.*, 2012; Peters and Turner, 2008). Since the development of DNA methodologies, the proliferation of studies

examining genetic diversity across a broad taxonomic spectrum, in turn has facilitated efforts to identify effects of selection on different gene regions (Tonteri *et al.*, 2010; Zueva *et al.*, 2014). The pressure to maintain increased loci diversity is associated with the requirement of individuals to identify and process a wide variety of infective insults (Lamaze *et al*., 2014a). Decreased diversity of MHC-associated loci raises infection potential across populations (Ejsmond and Radwan, 2011). The critical role played by the MHC in immune recognition means that the complex may be associated with a number of behavioral traits linked with the fitness of host populations (Consuegra and Garcia de Leaniz, 2008).

Two potential theories have been put forward to explain mechanisms that maintain MHC diversity. The first is denoted as "MHC heterozygote advantage" indicating heterozygous individuals display greater fitness due to a perceived ability to recognize a wider variety of parasite alleles increasing disease resistance (Doherty and Zinkernagel, 1975). Empirical data does give weight to the theory, within increased disease resistance of heterozygotes in murine models infected with *Salmonella enterica* (Penn *et al.*, 2002), Chinook salmon, *Oncorhynchus tsawytscha* (Walbaum, 1792) infected with haematopoietic necrosis virus (Evans and Neff, 2009) and Gila topminnows, *Poeciliopis occidentalis* (Baird and Girard, 1853) infected with *Gyrodactilus* (Hedrick *et al.*, 2001). The efficacy of the mechanism acting on wild populations has received skepticism associated with the significant increase in health between the healthiest homozygous individual and heterozygous (De Boer *et al.*, 2004). The second theory put forward to explain parasite-driven selection is the frequency-dependent selection hypothesis. Initially proposed as a purely mathematical model, the theory suggests a cyclical adaptation of host to most infective parasite allele followed by parasites adaptation to host selection (Slade and Macallum, 1992). The theory is backed by limited empirical studies deriving from gastrointestinal nematode (*Osteragia circumcincta*) infection in Scottish black face sheep, *Ovis aries* (Davies *et al.*, 2006)and MHC allele-specific resistance to *Aeromonas salmonicida* bacteria in captive bred at Atlantic salmon (*Salmo salar Authority*) (A. Langefors *et al*., 2001). Extrapolating the mechanism into wild populations presents temporal drawbacks. Alpinus *et al*. (1997) indicated that a

deleterious MHC allele would take 1000 generations to decrease from 1% to 0.1% within a population, with any long term study having to address convoluted timeframes to reach significant outcomes. So, although empirical data suggest that positive, parasite-driven selection underlies MHC variation the exact mechanism behind positive selection is still under debate (Sommer, 2005; Piertney and Oliver, 2006; Zhang *et al*., 2015a).

1.3.3 Functional variation of amino acid sequence within antigen binding site

Whilst MHC diversity goes part way towards maintaining population health, the successful infection of a parasite must require it to evade the MHC-dependent immune response. One factor that contributes to this is the amino acid sequence within ABS region, affecting final binding motif and rendering recognition ability of MHC void. Common mechanisms of parasite infection avoidance are; antigenic variation of parasite (Zambrano-Villa *et al.*, 2002), changes to the cell surface binding glycoprotein (Cowman and Crabb, 2006) and impaired T cell function (McSorley *et al.*, 2013). Within infective models of avian malaria the variation present within just one amino acid within the ABS allows of the parasite to evade the host's immune system (Radwan *et al.*, 2012). The nematode infection of Malagasy mouse lemurs (*Microcebus murinus*) denotes susceptibility to nematode infection with highly specific arginine placement within amino acid motifs within antigen binding sites (Schad *et al.*, 2005). To date molecular interactions between parasites and host immune cells are not well understood within wild populations. The current picture of host-parasite MHC relationships has relied heavily on studies done within controlled infection environments (Borghans *et al.*, 2004; GoüY De Bellocq *et al.*, 2008; Eizaguirre *et al.*, 2011; Dunn *et al.*, 2013). Investigating the intricacies of host-parasite relationships in the wild must encompass a multi-faceted approach. With the amino structure being so crucial to MHC efficacy, methodology design cannot focus only on MHC genetic diversity in response to parasite infection.

1.4 Host-parasite infection dynamics

1.4.1 Host-parasite molecular interactions

Freshwater trematodes frequently have complex life cycles utilizing multiple hosts within the freshwater environment before achieving definitive host infection (Criscione and Blouin, 2004). The parasite undergoes a significant transformation from free-living cercariae to blood-borne stage of infection within fish host (Curwen *et al.*, 2012.). Migration of the parasite through blood vessels requires combating a litany of host immunogenic responses. Defense against parasitic insult requires deployment of varied immunogenic countermeasures (Nordling, 1998; Sorrentino *et al*., 2002) Immunogenic response is broadly divided into two systems. The innate immune system is an immediate response to infection, utilizing direct binding of MHC molecule to pathogen initiating a T-cell immune cascade response (Maizels, 2009). The second form of immune reaction is the adaptive immune response involving pathogen-specific antibodies, which travel throughout the blood stream. These antibodies of the host recognize pathogenic insult in a highly specific manner to begin a complement-driven immune response (McSorley *et al.*, 2013). The capacity of freshwater trematodes to modulate and evade both aspects of the host immune response underpins invasion success and infection longevity within host populations.

1.4.2 Parasite tegument and surface antigens

The immune evasion of trematode parasites has been associated with the tegumental surface layer of the intermediate parasite stage (Coakley *et al.*, 2016). This tegumental surface is a syncytium that covers the entire worm bounded by a bilayer apical membrane, which interacts directly with host immune cells (Skelly and Wilson, 2006). The tegument is a folded unicellular membrane covered in micro-triches to increase overall surface area (Skelly and Wilson, 2006). It is, thus, a major canvas for scientists to isolate and investigate potentially immunomodulatory-important antigens used in trematode infection. One of these surface antigens frequently investigated is the Tetraspanins protein family. Tetraspanins are a large superfamily of evolutionarily-conserved cell surface antigens found throughout the natural world (Huang *et al.*, 2005).

A total of 37 Tetraspanin proteins were isolated within *Drosophila melanogaster* (Garcia-España *et al.*, 2008) implicated in a multitude of biological processes including; oocyte fertilization, infection susceptibility in mammalian parasites, cell-cell communication and immune cell interaction (Hemler, 2001). One of the most significant roles Tetraspanins have is in the action of cell adhesion in antibody binding processes, initiating changes in phosphorylation patterns within the cell-to-cell interface (Braschi *et al.*, 2006). To achieve immune-modulation tetraspanins proteins modulate crucial interactions of immune cells including adhesion, migration and synapse formation of membrane signaling complexes (Levy and Shoham, 2005a). The complex role that these proteins play in cell-cell processes is due to the structure and function of the proteins on the cell surface. The common structure of tetraspanins consists of four transmembrane domains. Transmembrane domains 1 and 2 flank a small extracellular loop with 3 and 4 flanking a larger extracellular loop. Within this second larger loop resides a distinct cysteine-cysteine-glycine (CCG) domain conserved across all tetraspanin proteins regardless of species (Hemler, 2003) (Figure 4A). This distinct tetraspanin structure is associated with specific functions; extracellular loops mediate specific protein-protein interactions with the intracellular regions associated with cytoskeletal strengthening of the tegument (Rubenstein, 1998). Specific immune functions affected by tetraspanins proteins have been narrowed down to; disruption of signaling marker for B cells (Tarrant *et al.*, 2003) and disruption of T-helper cell response (Mittelbrunn *et al.*, 2003). The disruption of T-helper cellular response has been implicated with the tetraspanins involvement in trematode infection. Decline in T-cell titer during blood stage infection of Schistosoma *mansoni* (Lewis and Wilson, 1981) and decreased T-cell recognition in *Fasciola hepatica* (Robinson *et al.*, 2009). Through its direct involvement with the immune system tetraspanins proteins have been highlighted as potential vaccine targets for Schistosomiasis. Although investigations are still in their infancy, TSP-2 knockdown studies have seen an increased egg clearance rate within murine models (Tran *et al.*, 2006).

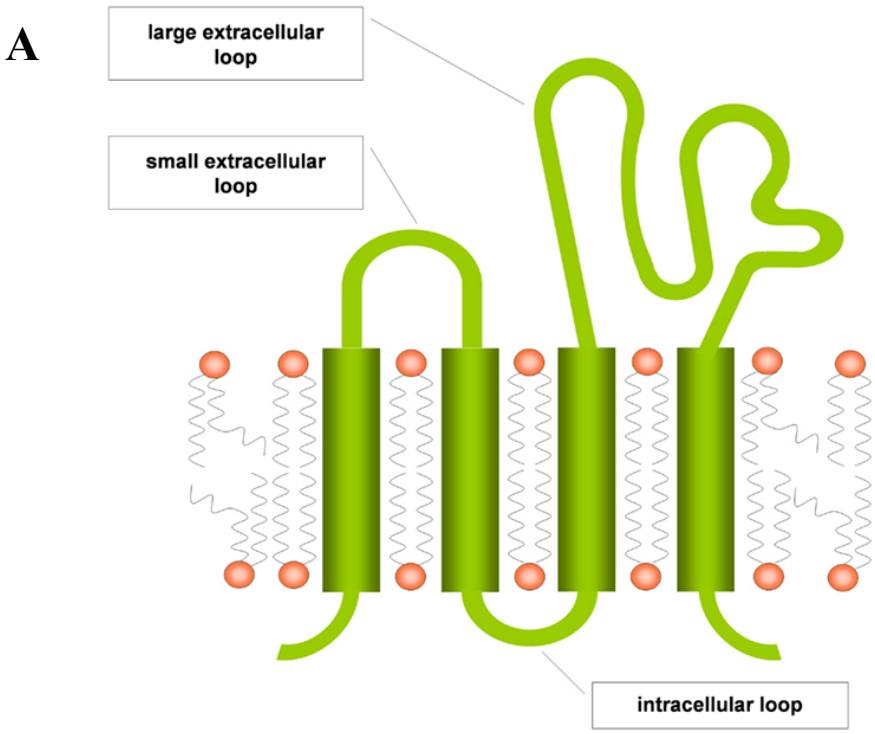


Figure 4; Two potentially immunomodulatory proteins utilized in freshwater trematode invasion. A, Tetraspanin intracellular protein with distinctive double extracellular loop (Kaesler et al., 2012). B, Venom allergen-like protein, one of many proteins released within blood-borne infective stage that down regulates immune cell action of host (Kelleher et al., 2014)

There are further proteins released as an immunomodulatory cocktail during parasite blood stage infection. Included among these is the Venom-Allergen Like (VAL) family of proteins(Chalmers and Hoffmann, 2012). VAL proteins are a large superfamily of proteins all containing a conserved 3-layer α-β-α sandwich tertiary structure (Hewitson *et al.*, 2011) (Figure 4B). Using a combination of genomic and proteonomic analysis the VAL family can be divided into two distinct groups, Group 1 VALs (VAL1, 2,3,4,5,7,8,9,10,12,14,15,18,19,20,21,22,23,24,25,26,27,28) are thought of as signal peptides with 3 conserved disulphide bonds. Group 2 VALs (VAL6, 11 13, 16, 17) do not possess these bond groups (Curwen *et al.*, 2006). There is growing evidence to suggest that group 1 VAL proteins are an excretory product in trematode species as disulphide bond indicates an increased protein-protein binding potential. Investigations have discovered group 1 proteins within *S. mansoni, S. japonicum* (Curwen *et al.*, 2006*)* and *O. viverrini* (Young *et al.*, 2014). Immunomodulatory actions were observed causing a depleted IgG response during *S. mansoni* murine infections (Pleass *et al.*, 2000) and reduced T-helper efficacy during host migration (Jenkins and Mountford, 2005). Information regarding group 2 proteins is sparse; the one exception is VAL 6 protein being extensively expressed throughout host invasion and infection of *S. mansoni* (Hansell *et al.*, 2008). Due to the large variety of VAL proteins isolated, studies into the efficacy of using them as drug or vaccine targets remain scarce (Hemler, 2003).

Freshwater trematodes have adapted varied immuno-evasive tools to match the multi-armed host defense response. Genomic-based research has allowed for the annotation of specific proteins that affect the effectiveness of infection. With both of the above proteins playing crucial roles in infection, it is natural that they should be highlighted as potential targets for vaccine design. Although genomic-based investigations have been based on medically-relevant parasites *Schistosoma* and *Fasciola*, focusing future research on aquatic trematodes such as *Diplostomum* may provide targets for vaccines that can be used in aquaculture practices, decreasing the overreliance on antihelminthic drugs in the fish farming industry.

1.5 Rationale and aims

1.5.1 Study area

Populations of *S. trutta* (brown trout) have existed in self-sustaining populations in the northern Highlands of Scotland since the retreat of last Scottish ice sheet (Late Devensian; 17,000 – 13,000 years ago) (McKeown *et al.*, 2010). They present as anadromous (sea trout), lacustrine (brown trout) and carnivorous (ferox trout) morphs. The lacustrine brown trout is known to populate the many hill lochs, estuaries and rivers that populate the Scottish Highland environment (Neil, 2013). Demonstrated in numerous fish health screens, UK brown trout exhibit high levels of parasitism and are frequently infected by various species of cestode, nematode and trematode (Byrne *et al.*, 2002; Hartvigsen and Kennedy, 1993; Molloy *et al.*, 2014). Brown trout inhabiting the Gairloch region of Scotland populate an assortment of isolated freshwater hill lochs and sub-alpine streams in this region. Populations of trout in the region have not undergone substantial stocking of non-native species, unlike many salmonids populations across Eurasia (Schenekar and Weiss, 2017).With sub-populations of native brown trout existing alongside parasite

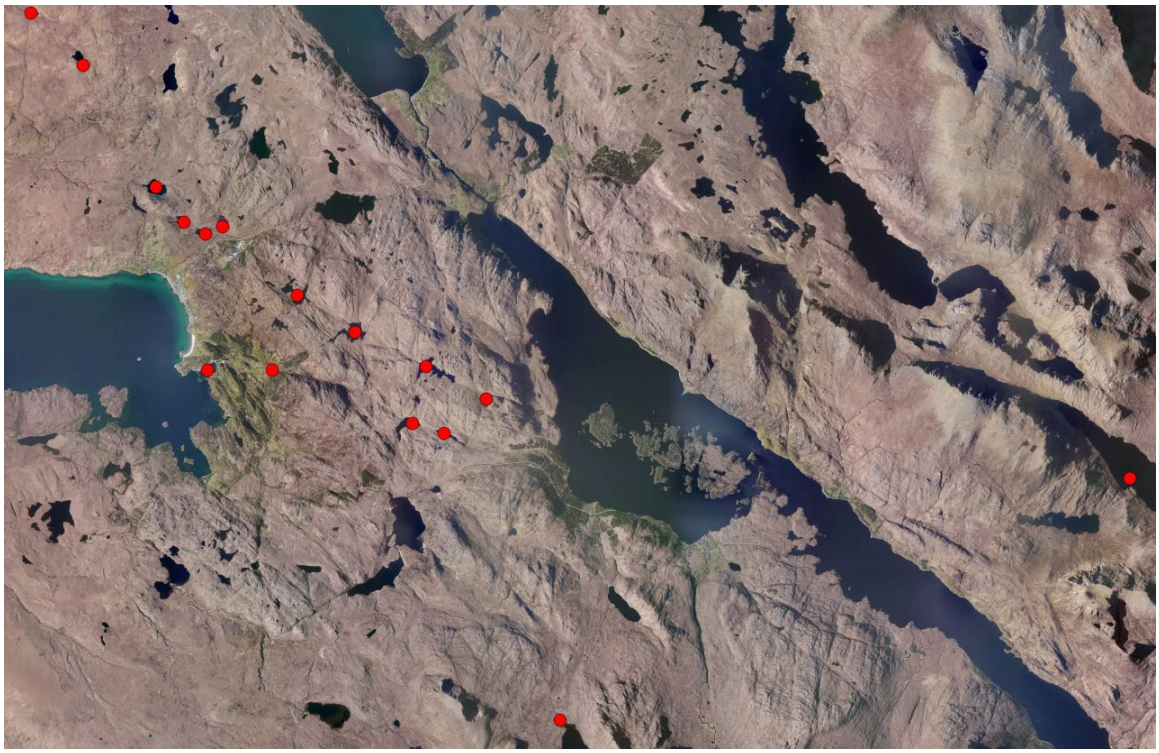


Figure 5; Gairloch hill loch system within the northwest Scottish Highlands. The study area presented numerous elevated lochs and sub-alpine burns all inhabited with self-sustaining brown trout populations. Map was created using ggplot R packages (Wickham, 2016).

Fauna in isolation, it makes the region a model area to study the long term effects of parasite presence on trout immunogenetics (Figure 5).

As highlighted within this review, genomic data (remember data is plural) surrounding host-parasite relationships within the aquatic environment are sorely lacking. Salmonid aquaculture has eased some population pressure on wild fish stocks, however the continued reliance on pharmalogical intervention to decrease parasite infection (and other factors) has led to a severe negative impact on wild populations of salmon and trout. The understanding of the infection dynamics and population genetics of aquatic parasites could aid in future control methods within aquaculture industry and conservation practices. This study will seek to highlight the importance of genetic species identification and utilizing genomic techniques to understanding molecular factors associated with parasite resistance and susceptibility within populations of salmonids. The overarching aim will be achieve through the completion of these key objectives:

Objective 1: To perform a full parasite screen of brown trout subpopulations across the Gairloch hill Loch system within varied locations and water bodies.

Objective 2: To carry out characterization key parasites species infecting brown trout across the Gairloch system. Species level identification will be performed utilizing molecular markers associated with species delineation to provide accurate depiction of species infecting subpopulations of brown trout.

Objective 3: To carry out a population-wide analysis of the antigen binding site in MHC class II resulting data with differential parasite infection between trout populations

Objective 4: To carry out a genomic study of antigenic properties of *Diplostomum* spp. affecting infective success across Gairloch system. This process will include the genomic annotation of key antigens associated with immunomodulation of the parasite.

2 Survey of parasites infecting wild brown trout in a highland hill loch system

2.1 Introduction

Brown trout is an indigenous salmonid species in Europe, North Africa and western Asia (L Bernatchez, 2001), it exists in three phenotypic morphs; the anadromous sea trout (*S. trutta trutta*) and lacustrine form (*S. trutta fario*) (A Klemetsen *et al*., 2003). They live in brackish estuaries, streams and lakes and spawn in fast-flowing streams (Cucherousset *et al*., 2005). To inhabit a variety of environments residential brown trout are eclectic predators. Their main food source is littoral zoo benthos, however populations adapt to particular food item when specific to habitat and at times can become *largely* piscivorous/carnivorous (L'Abée-Lund *et al.*, 1992). Diet changes within a population are influenced by development rate and size of the fish within the habitat. Smaller brown trout are primarily surface feeders near the shore, with larger fish more inclined to exploit pelagic waters offshore, feeding on zooplankton, copepods and small fish. Trout food choice can increases likelihood of parasitic infection due to intermediate host consumption (Byrne *et al.*, 2002).

2.1.1 Brown trout health in Scotland

The original populations of brown trout in Scotland occupied the isolated lochs, streams, and highland lochans immediately after the retreat of the European/Scandinavian ice sheet approximately 12,000 years ago (L Bernatchez, 2001). Salmonid species in Scotland have received recent attention due to the substantial decline of anadromous and estuarine populations along the west coast of Scotland over the last 20 years (Bjørn *et al.*, 2001; Krkosek *et al.*, 2005). The dramatic population decline across anadromous populations is a likely consequence of expansion of offshore fish farming across the west coast of Scotland the, with the concomitant increase in *Lepeophtherirus salmonis* (salmon lice) infecting annual sea trout runs (MacKenzie *et al*., 1998). Equally, the occurrence of parasite infection within freshwater fish has received considerable attention over the last

30 years (Nachev and Sures, 2016; Sheath *et al.*, 2015; Siwertsson *et al.*, 2016). There are two schools of thought when examining the occurrence and predictability of parasite fauna within resident fish species. One theory explains that the infection of parasites within freshwater species is wholly stochastic with chance colonization of parasite host species (Kennedy 1978, Valtonen *et al*., 2001). In contrast, to this theory is the supposition that rare species of parasite exhibit predictability in their occurrence, which is, facilitated by specific environmental factors. Such a case might be the Norwegian brown trout populations in isolated lakes where individual rare parasite species occur in specific locations, owning to snail host presence (Hartvigsen and Halvorsen, 1993; Byrne *et al.*, 2000). Populations of brown trout frequently act as the intermediate hosts of numerous helminth species (Hartvigsen and Kennedy, 1993), are infected by trematodes (Moody and Gaten, 1982), Cestodes (Dorucu *et al.*, 1995), Nematodes (Xiong *et al.*, 2013) and Acanthacephala (Nabi *et al.*, 2016). The complexity of many parasite lifecycles means that transmission success rate depends on host availability and size of water body (Kennedy, 1985). Summer months represent a time period of higher infection rate for many parasite species, warmer water allowing extended parasite survival times in free living forms. Within higher latitude regions the reduction in warm month duration constrains the infective capability of freshwater helminthes (Hakalahti *et al.*, 2006; Marcogliese, 2008; Zbikowska and Nowak, 2009). Intestinal parasites make up the majority of infective load in trout populations. Cestode (*Diphyllobothrium spp.*) and Nematode (*Eustonglyides spp.*) infections have significant pathogenic impact in fry but minimal pathogenic effects on adult brown trout (Dorucu *et al*., 1995; Riitta Rahkonen *et al*., 1996) . Extra-intestinal helminth infection sites in brown trout include; swim bladder (*Cystidicola farionis*) (Siwertsson *et al.*, 2016), eyeball (*Diplostomum spp.* and *Tylodelphus*) (Stumbo and Poulin, 2016), gills (*Ergasilus sieboldi*) and mucosal epithelium (*Pisciola, Hemiclepsis and Gyrodactilus*) (Buchmann and Uldal, 1997; Loch and Faisal, 2014). Freshwater Digenea represent an important parasitic agent in trout, with indirect pathogenic effects on the host. The eye flukes *Diplostomum* and *Tylodelphus* metacercarial infective stage may blind the fish in heavy infection loads, making the fish vulnerable to increased predation, and have been associated with

increased fish deaths in farmed (Larsen *et al.*, 2005) and wild populations (Blasco-Costa *et al*., 2014b).

2.1.2 Parasite identification

Delimiting species of parasite is often difficult due to their limited morphological characteristics. Additionally, many parasites recovered from host samples may be larvae or eggs. Thus genetic identification of parasite species is quickly becoming the most utilized application in species identification in parasitology. Parasite screening of host samples involves thorough dissection procedures and the removal of parasites residing within host. Identification of fish parasites is important for the monitoring of disease transmission and fish health screening within a freshwater system; it is particularly important because many parasites are pathogenic to humans (Wicht *et al*., 2010; Fang *et al*., 2015). Many fish parasites share morphological characteristics, making accurate identification of a large dataset difficult and exacerbating public health issues associated with, and requiring accurate identification of, such parasites.

DNA barcoding offers a widely regarded method to provide initial taxonomic identification of parasite species compared to previously used morphometric practices (Hebert *et al.*, 2003). The amplification of the cytochrome oxidase gene via parasite tissue sample offers a DNA-based 'barcode' of the parasite to provide an initial delineation of parasite species. This is particularly helpful in ascertaining parasite fauna of freshwater fish species (de León and Choudhury, 2010; Xiong *et al*., 2013) with morphometric analysis provides a time consuming and inaccurate way of identifying parasites of similar morphology (Hajibabaei *et al*., 2007; I. Blasco-Costa *et al*., 2016). One particular strength of DNA barcoding techniques is its ability to process a large amount of samples relatively quickly, allowing for large-scale identification of screened parasite data from host populations (I. Blasco-Costa *et al*., 2016; de León and Choudhury, 2010; Hajibabaei *et al*., 2007). The identification of the gene sequence is performed when compared against similar sequences using blastn algorithm (Ye *et al*., 2006). The algorithm uses the heuristic search criteria to match sections sequences in a stepwise manner to provide similar sequences revealing species identity (Ye *et al.*, 2006).

Parasites are an important influence upon wild brown trout, both in terms of their conservation in the wild and in their role as proxies for studying socioeconomically important salmonids in aquaculture. This, coupled with the rising incidence of fish parasites within human populations (Kuchta *et al*., 2013) means that the recently-acquired ability to carry out molecular characterization of the incidence and interactions of brown trout and their parasites at the population level offers opportunities to examine questions that have been hitherto difficult to address. This study takes advantage of this opportunity.

2.2 Study Area

The study area consists of 12 lakes (lochs) in the Gairloch hill loch system on the north west coast of Scotland (Figure 1). The entire highland region of Scotland underwent dramatic deforestation in the last 300 years; this removed much of the native plant fauna from the highlands and Gairloch in particular (Dickson, 1977). Along with the removal of native forests, large predator removal has also occurred in the region, with the complete extinction of large mammal predators; the Eurasian bear (*Ursus arctos arctos*) and the Eurasian Wolf (*Canis lupus lupus*) (Nilsen *et al.*, 2007) and a drastic reduction in birds of prey species; the Osprey (*Pandion haliaetus*) and White-tailed eagle (*Haliaeetus albicilla*) (Whitfield *et al.*, 2004). Deforestation has proven to be a major factor in the reduction of nutrients in the environment, leading to species extinction and has been exacerbated by removal of apex predators from the system, drastically reducing the motility of nutrients through local food webs and local habitat (Innes, 1983). The investigation of parasitic species infecting the brown trout in the area would provide a database of parasites infecting an important salmonids species in the UK and monitor the presence of any rare species, or particularly virulent species of parasites. The parasite screen would also provide population and community dynamics of species across an oligotrophic loch system with a diminished definitive host presence.

2.3 Materials and Methods

2.3 Sample protocol

The sampling sites of wild brown trout are shown in Fig 1. A total of 210 individuals were sampled across the Gairloch hill lochs within 14 populations (Fig 1). Anglers using hook and line through the years 2012-2015 donated the collections of whole trout samples. Collection periods were during high parasite months of July – August; this regime was repeated during 2012, 2013, 2014 and 2015.

Fish were caught and dissected fresh within 1-3 hours after capture, with dry weight and fork length of fish taken for each fish. All fish underwent complete necropsy according to fish health screen protocol (Environment agency, 2011). The parasitological investigation of the trout began with an external inspection of the skin, fin, gills, and mouth cavity for the detection of possible ecto-parasites. Once exterior is thoroughly examined blood was extracted from heart to examine potential protozoa infection. The body cavity was opened, particularly the liver, gonads, kidneys, and external surface of the digestive tract. Stomach and intestines was opened and potential infecting parasites removed. The parasites were isolated, removed carefully and prepared for identification mostly to the species level. Investigation of removed parasites was performed using comprehensive microscopy analysis for protozoan and metazoan parasites. All parasites were preserved in 70% alcohol and stored for later identification. Individual parasites were stored in individual 1.5ml micro centrifuge tubes to reduce the chance of cross-species molecular contamination.

	Map Sq.	North	East	Elevation (m)	Weight	Length
Lochen nam Breac	NG	785	813	79.6	38.4±15.1	149.5±20.6
Loch Feur	NG	783	806	72.3	40.6±17.3	144.6±26.3
Loch Coire na-h-Arigh	NG	784	804	79.9	43.3±17.3	152.8±25
Loch Gharbe Doire	NG	792	796	84.2	68±59.9	175.9±52.9
Loch Airigh Mi-Craidh	NG	827	768	219.8	63.1±9.6	144.8±15.1
Loch Airigh a'Phuill	NG	846	755	232.9	65.1±30.1	175.3±33.2
Loch na Feithe Mugaig	NG	745	856	302.3	40.7±17.3	161.4±29.2
Loch Doire na h-Airigh	NG	741	874	271.2	90.7±29.5	206±24.3
Loch nam Buainichean	NG	736	854	203.1	63.5±36.3	171.4±35.3
Lochan Dubh nam Bias	NG	736	859	218.8	214.5±38.1	128.8±64.2
Loch na h-Oidche	NG	668	884	387.6	154.8±19.1	47.2±13.5
Lochan Fada	NG	685	648	310.1	215.5±27.1	207±44.5
Loch a' Mhadaidh	NG	788	799	168	213.3±23.2	177±24.6
Alt a' Glinne	NG	843	753	229	75.5±9.7	119±44.7
Lochan Sgeireich	NG	786	800	127	155.5±29.5	162±44.8
Flowerdale estuary	NG	843	748	6	215.5±27.5	136±74.9

Table 1; Sample site location properties. Coordinates of sample sites in congruence with Ordinance survey map coordinates. Weight and Length denote standard deviation of average dry weight and fork length of brown trout caught in each location

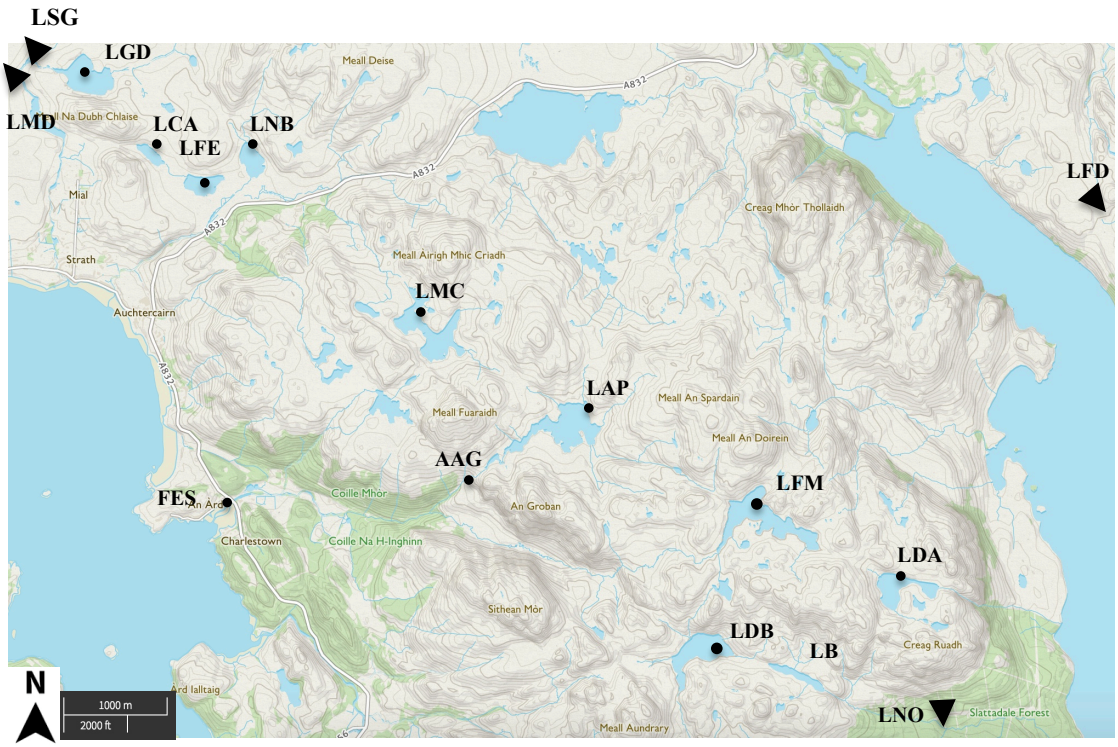


Figure 6: Map of sampling regions within Gairloch system abbreviations are used for specific loch sub-populations: LNO; Loch na h-Oidche, LFM; Loch na Feithe Mugaig, LDA; Loch Doire na h-Arigh, LAP; Loch Arigh a'Phail, LMC; Loch Arigh Mhic Criadh, LFE; Loch Feur, LNB; Lochen nam Breac, LCA; Loch Coire na-h-Arigh, LGD: Loch Gharbe Doire and LFD; Loch Fada. Triangles denote direction of loch not fitting on map (osmaps online, 2018)

2.3.1 Molecular identification

Total genomic DNA was extracted using protocol of Qiagen™ Blood and Tissue extraction kit. Primers were designed using previously published species-specific primers and using primer-BLAST software from NCBI (www.ncbi.nlm.nih.gov/tools/primer-blast/). Primer sequences, fragment size and published reference are contained on table 2 primer design table. Initial species identification was performed using light microscopy to sort species before PCR procedure.

PCR reactions were performed for each gene fragment using 12.5 µl hot start taq polymerase; Thermo–Start^{R}PCR master mix (0.625 Units of Taq DNA polymerase, 1X reaction buffer, 0.2 mM of each dNTP and 1.5 mM $MgCl_2$) Reactions were performed using a Techne Prime 96 well thermal cycler (Techne™) and 5 µl of each amplicon was visualized in 1 % agarose gels stained with SafeView nucleic acid stain (nbsbiologicals™) under UV using Licorgel documentation system. The remaining 20 µl PCR products were sequenced using Sanger sequencing protocol at the DNA sequencing facility of the Natural History Museum, London, using the PCR primers with Fluorescent Dye Terminator Sequencing Kits (Applied Biosystems™); sequencing reactions were run on an Applied Biosystems™ 3730XL automated sequencer.

Identification of amplified fragments was analyzed using nucleotide blast database as part of the NCBI online suite of nucleotide analysis software (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The setup up of blast analysis was to last sequences against "Nucleotide collection (nr/nt)" ncbi database, utilizing blastn algorithm To ensure the most thorough analysis of similarity between acquired amplicons and known nucleotide database collection.

Species	Primer sequence F'	Primer sequence R'	amplicon size	ref
Diphyllobothrium dendriticum	5'-TTT TTG GCC ACC CCG AAG TAT AT-3'	5'-TAG TGA CAT TAC ATA GTG GAA GTG-3'	910bp	Wicht et al., 2007
Eubothrium crassum	5'-TTT TTG GCC ACC CCG AAG TAT AT-3'	5'-TAG TGA CAT TAC ATA GTG GAA GTG-3'	613bp	
Eubothrium salvenini	5'-TTT TTG GCC ACC CCG AAG TAT AT-3'	5'-TAG TGA CAT TAC ATA GTG GAA GTG-3'	601bp	
Diplostomum baeri	5'-TGT AAA ACG ACG GCC AGT-3'	5'-CAG GAA ACA GCT ATG ACT-3'	384bp	Moszczynska et al ., 2009
Neoechinorhynchus saginatus	5'- GTCGTAACAAGGTTTCCGT-3'	5'-ACCCGCTGAATTTAAGCATA-3'	987bp	Luton et al., 1992
Eustrongiyides spp	5' -ACNACRTARTANGTRTCRTG-3'	5' -TGRTTYTTYGGNCAYCC-3'	419bp	Hafner et al., 1994

Table 2: Table of cox1 primers designed to identify parasites infecting populations within the Gairloch system. Newly designed primers are denoted in yellow boxes.

2.3.2 Data analysis

Indices of the prevalence and mean abundance were calculated for each parasite species. The normality of the data series was tested using Kolmogorov-Smirnov test as part of Minitab v13 statistical analysis program. The Core-satellite species hypothesis is one that explains the occupancy patterns of species across a landscape. Core species are typically broadly distributed and abundant within the region studies. Satellite species are less broadly distributed and typically denoted as rare within the study region (Hanski, 1982; Gibson *et al.,* 2005). Analysis of parasite community and the investigation of core and satellite species involved the measurement of the prevalence within individual population in comparison to entire dataset; >60% = core species, 41-60% = secondary species, 6 – 40% = satellite species and <6% = rare species. The handling of the number of parasites and distribution was split into three different categories:

Prevalence: it is intended as a descriptive statistic for the presence or absence of parasite species infection. It is commonly displayed as a percentage when used descriptively (Burns *et al*., 2007)

$$\textbf{Prevalence} = \frac{\text{Total number infected}}{\text{Total number of fish}} \times 100$$

Abundance: Abundance is a number denoted as the number of individuals that are infected by the particular parasite within the sampled population (Burns *et al*, 2007).

$$\textbf{Abundance} = \frac{\text{Total number of parasites recovered within sub-population}}{\text{Total number of fish examined within the sub-population}}$$

Raw parasite number: This is denoted as the number of parasites recovered within the sample subpopulation within mean incidence of the species across the sub-population denoted the incidence of individual parasite (95% confidence interval) (Burns *et al*., 2007).

2.4 Results

The total parasite fauna consisted of 6 taxa including three species of cestodes *Diphyllobothrium dendriticum* (Nitzsch, 1824)*, Eubothrium crassum* (Bloch, 1779)*, Eubothrium salvelnini* (Schrank, 1790) one species of digenea (*Diplostomum baeri.*), one species of nematode (*Eustronglyides* spp. (Jagerskiold, 1909)) and one species of acanthacephalan (*Neoenchinorhynchus saginatus* (Stiles and Hassall, 1905)). Resulting blastn analysis of sequences revealed high similarity of cestode species *D. dendriticum* (ident 98%, E value =0.0), *E. crassum* (ident 97%, E value=0.0), *E. salvenini* (ident 97%, E value=0.0). The lowest blastn scores were within digenea *D. baeri* (ident 93%, E value=0.0) and acathocephala *N. saginatus* (ident 89%, E value 0.0) (Table 3).

The E value is fully named the 'Expect value', and is a parameter that describes the number of hits that is expected to match per chance when searching the Blast database. It decreases exponentially as the similarity score increases, essentially explaining the amount of 'background noise' hits that occur when the blast algorithm matches the sequence. The closer the E value is to zero, the more significant the match is with zero background matches occurring as part of the output. As the entire test of E values (Table

3) were 0.0 indicating all retrieved sequences were significantly matched to the corresponding cox1 sequence of Blast match.

Species name	Query cover	Identical %	E value
Diphyllobothrium dendriticum	100%	98	0
Eubothrium crassum	100%	97	0
Eubothrium salvenini	100%	97	0
Diplostomum baeri	98%	93	0
Neoechinorhynchus saginatus	90%	89	0
Eustronglyides spp	99%	97	0

Table 2; Results table from blastn algorithm search on NCBI genbank database. Matching identical % includes average identical value across populations of parasite with standard deviation of mean identical % score.

Species presence was divided between localities; 4 sites displayed cestode infection Lochan nam Breac, Loch Feur, Loch Coire na h-Airigh and Loch Gharbe Doire (fig 4). The only taxon that infected all sites was the digenean, *Diplostomum baeri.* with a prevalence of 81% across the study area. The nematode taxa *Eustronglyides* spp. occurs only in two lochs. *Diplostomum baeri* dominated the brown trout parasite community and represented the core species both overall and in every locality. Brown trout parasite infracommunities consisted of 1-6 species with 12% being completely uninfected across the sample site (Hanski, 1982; Gibson *et al.,* 2005). Single species infracommunity infection prevailed across the entire sample site with of all fish being infected by one

species (98% being *Diplostomum baeri*). Multiple species infection occurred predominantly within the low-lying lochs; Lochen nam Breac, Loch Feur, Loch Coire na-h-Arigh and Loch Gharbe Doire (fig 5), with all trout (47) sampled showing co-infection by more than one species of parasite, 60% being infected by 3 species and 32% being infected by 4 or more species. As the most dominant parasite *Diplostomum baeri* indicated an increased tendency to be associated with infra-community infection in all populations sampled compared to other infective species.

The community classification (Hanski, 1982; Gibson *et al.,* 2005) of the parasites in the system saw all cestode and nematode species being classified as secondary species (*D. dendriticum* 2.21%, *E. crassum* 9.75%, *E. salveini* 3.23% and *Eustronglyides* spp. 7.68%), with Acanthacephala being considered rare species (*N. saginatus* 0.69%) and the core species across all populations being *Diplostomum spp.* (76.41%) (fig 5). The parasite taxa that showed the highest mean abundance was *Diplostomum baeri,* with consistent high infection numbers of host across the entire dataset. Three sub-populations indicated increased abundance of an alternative species to *D. baeri* ; Loch Coire na-h Arigh (*E. crissum*), Loch Feithe Mugaig (*Eustronglyides spp.*) and the highland burn Alt a'Glinne (*E. crassum*). *Eubothrium crassum* was the second most abundant parasite with Loch Coire na-h Arigh showing the highest abundance, with 10 fish from the sample population being infected with the parasite. Alt a'Glinne demonstrated the highest degree of *E. crassum* prevalence at 83% the only sub-population sampled that had did not have *D. baeri* as the most prevalent parasite. *Eustronglyides* only was present within two sub-populations sampled however within the Loch Feithe Mugaig population, the parasite showed similar abundance and prevalence to the most dominant parasite within the system, *D. baeri*.

Species	LNB	LFE	LCA	LGD	LAM	LAP	LFM	LDA	LB	LDB	LNO	LF	LMD	LSG	AAG	FES
Neoechinorhychus sp.																
Prevalence (%)	16.00%	5.04%	1.29%	/	/	/	/	/	/	/	/	/	/	/	/	/
Abundance	0.4	0.16	0.11	/	/	/	/	/	/	/	/	/	/	/	/	/
raw infection number (per individual)	3 (0.23±0.8)	6 (0.55±1.2)	2 (0.16±0.57)	/	/	/	/	/	/	/	/	/	/	/	/	/
Diphyllobothrium dendriticum																
Prevalence (%)	8.00%	1.68%	5.16%	6.71%	/	/	/	/	/	/	/	/	/	/	/	/
Abundance	1.9	0.25	0.32	1.33	/	/	/	/	/	/	/	/	/	/	/	/
raw infection number (per individual)	15 (1.15±0.9)	5 (1.55±0.9)	8 (0.66±0.7)	3 (0.2±0.6)	/	/	/	/	/	/	/	/	/	/	/	/
Eubothrium salvenini																
Prevalence (%)	/	5.04%	10.96%	/	/	/	/	/	/	/	/	/	/	/	/	/
Abundance	/	0.30	5.32	/	/	/	/	/	/	/	/	/	/	/	/	/
raw infection number (per individual)	/	3 (0.55±0.24)	17 (1.41±1.1)	/	/	/	/	/	/	/	/	/	/	/	/	/
Eubothrium crassum																
Prevalence (%)	29.18%	21.84%	30.96%	1.34%	/	/	1.77%	/	/	/	/	/	/	/	83.00%	/
Abundance	8.5	4.66	10.55	1.55	/	/	0.44	/	/	/	/	/	/	/	9.66	/
raw infection number (per individual)	54 (4.15±0.6)	17 (2.15±1.74)	48 (4±1.1)	1 (0.15±0.04)	/	/	82 (7.45±1.8)	/	/	/	/	/	/	/	87(8.8±7.8)	/
Eustronglyides spp.																
Prevalence (%)	/	/	/	/	/	22.34%	48.52%	/	/	/	/	/	/	/	/	/
Abundance	/	/	/	/	/	1.65	8.66	/	/	/	/	/	/	/	/	/
raw infection number (per individual)	/	/	/	/	/	42 (2.1±5.3)	3 (0.27±0.9)	/	/	/	/	/	/	/	/	/
Diplostomum baeri																
Prevalence (%)	45.00%	66.38%	51.61%	91.94%	100.00%	72.87%	49.70%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	17.00%	100.00%
Abundance	17.4	6.50	9.11	15.77	8.25	10.20	3.22	12.00	3.50	7.44	10.50	22.00	15.40	8.40	1.66	11.77
raw infection number (per individual)	85 (6.5±0.8)	79 (7.18±0.7)	80 (6.66±0.5)	137 (13.7±7.5)	153 (8.5±7.2)	166 (9.76±7.8)	84 (7.63±1.2)	77 (7±4.6)	42 (3.81±0.9)	34 (2.83±1.6)	110 (8.46±7.5)	158 (12.15±11.8)	110 (11.23±9.9)	53 (5.1±4.3)	17(2.8±1.2)	103 (9.2±8.1)

Table 3; Parasite number infecting populations in separate lochs. Species are denoted via genus name and species. Number of parasites indicated via total amount of parasites found in loch populations with mean and standard deviation of parasites denoted in brackets beneath.

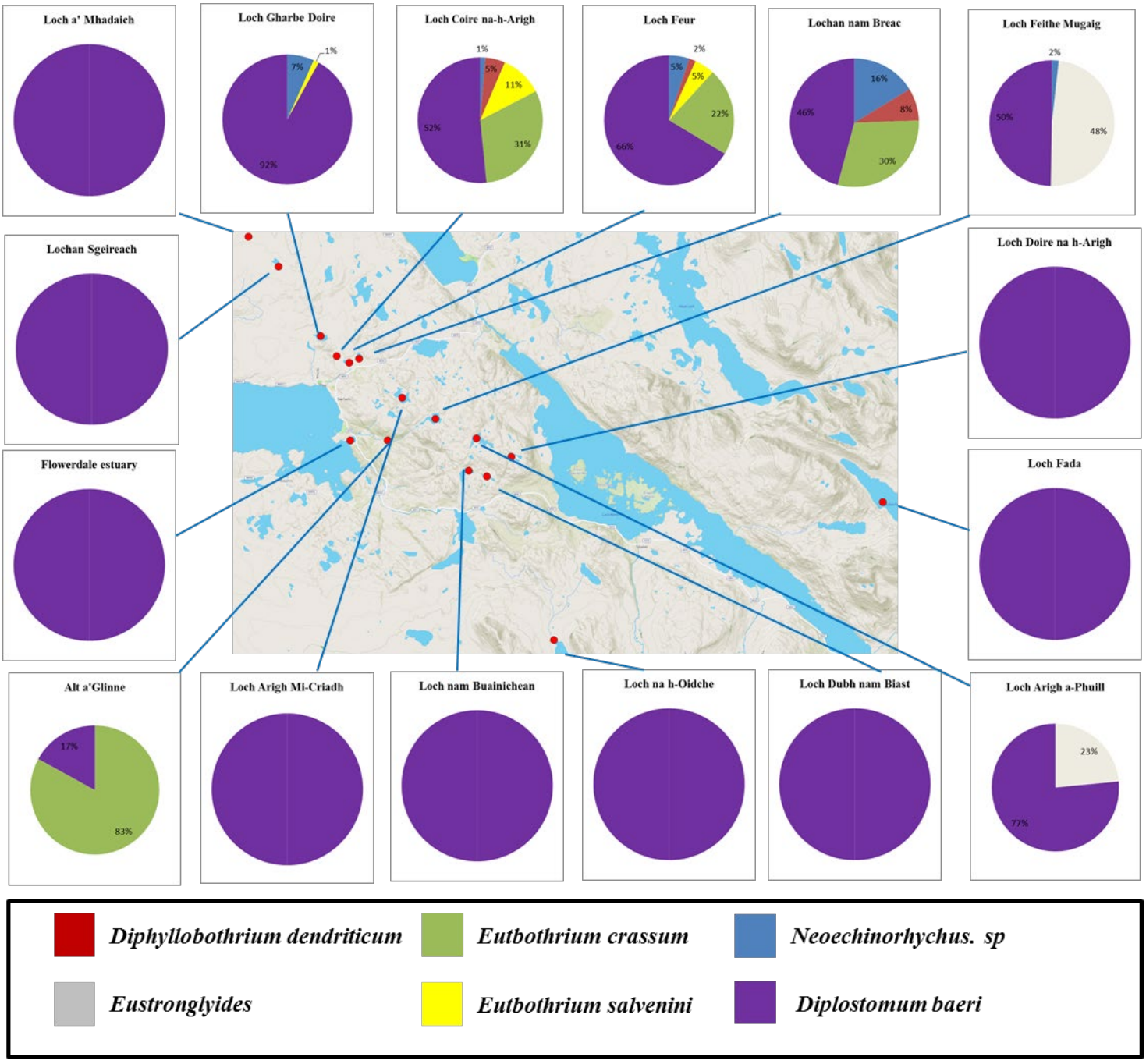


Figure 7: Parasite prevalence mapped across the system using percentage prevalence of each of the six species mapped alongside the sample sites

2.5 Discussion

Molecular identification of parasites is increasingly being used for species determination. An open-source online platform using blastN analysis makes rapid species identification easy to perform, matching input sequence with previously published marker sequences within library. In this case, identification of parasites within the Gairloch Hill Lochs formed a preliminary analysis carried out in order to gain an overview of parasites infecting brown trout in the system.

All parasites infecting brown trout within the Gairloch hill lochs were representative of European fauna and no invasive parasite taxa were recorded. In total six parasite species (table 2) were characterized in the study, greatly extending the knowledge of parasite infection in lacustrine brown trout in Scotland. Previously the only been extensively studied morph was the anadromous form (*S. trutta trutta*) (Dorucu *et al.*, 1995; MacKenzie *et al*., 1998). Results from anadromous parasite screen revealed an increase in parasite infection diversity. Anadromous forms saw infection from all parasites found in this freshwater screen, however the increased diversity of cestode infection; *Cyathocephalus truncates* (Pallas, 1781) and *Diphyllobothrium ditreum* (Creplin, 1825) and nematode infection (*Capillaria salvelnini* (Polyanskii, 1952), *Raphidascaris acus* (Bloch, 1779) and *Cystidicola farionis* (Fischer, 1798)) was evident in sea trout parasite screens (Dorucu *et al*., 1995). Substantial copepod infective presence (*Lepeophtheirus salmonis* (Kroyer, 1837) and *Caligus elongatus* (Von Nordmann, 1832)) has also been indicated in numerous sea trout parasite screens on the west coast of Scotland (Johnson and Albright 1991; Piasecki and MacKinnon, 1995; Mackenzie *et al*., 1998). Substantial copepod infection has not as of yet been discovered within lacustrine brown trout populations and has not been the case in this investigation. Copepod presence would have greater infective potential on brown trout residing within estuarine populations of residential brown trout. Freshwater localities are often characterized by the differences in parasite communities infecting fish host that inhabit them (Hartvigsen and Kennedy, 1993; Molloy *et al*., 1995; Valtonen *et al*., 2001; Nelson and Dick, 2002). In many cases the parasites infective capability is limited by suitable host presence. This is particularly true in the case of parasites with more complex life cycles where host density is crucial in

successful transmission of the parasite's infective stage. Environmental factors play a key role with the presence or absence of host species being crucial for a parasite species to maintain infection in a freshwater body.

Parasite abundance as a quantitative parameter of parasite infections is considered to be multifactorial in nature, as a function of parasite spatial variation it can be a function of a multitude of abiotic and biotic factors (Poulin, 2007). The aggregation of a parasite within a system can be considered as a highly variable species attribute, which is dependent on the intrinsic to the biology of the parasite, host species and environment (Arneberg, 2002; Poulin and Mouillot, 2003; Kutz et al., 2009; Beldomenico and Begon, 2010). The variability of the parasite abundance and prevalence is important from an ecological point of view, a potential regulator of parasite populations and influencer of spatial structures (Morand and Guegan, 2000; Smith, 2001). It also demonstrates importance from an epidemiological point of view, where the importance of parasite abundance in the spread of parasitic diseases amongst host populations is a stressed subject

In this study *Diplostomum baeri* indicates a higher infective level than other parasites species, with trout sampled from all lochs indicating eye fluke infection to varying degrees. Within Scotland, the parasite exists at high levels within farmed populations (Stables and Chappell, 1986a) with this being the first study to see wide-spread infections in the wild. Piscivourous bird host has been isolated as a principle cause for broad-scale geographical infection presence mirroring studies in of the St Lawrence river in Canada (Locke *et al*., 2010). For parasites with complex life cycles that utilize sequentially more than one intermediate host species an overlap of host species is vital for successful transmission (Wilbur, 1980; Prugnolle *et al*., 2005; Louhi *et al*., 2010). The distribution of the parasite across the Gairloch environment indicates sufficient snail host density to perpetuate infection level. One factor in regarding high infection levels is annual accumulation of parasites during the high infectivity summer months (Hakalahti *et al*., 2006). Sub-populations with higher infection rates may correlate with an increased age of individuals within the population, accurate aging comparisons between subpopulations

are needed to ascertain correlation between fish age with increased presence of *Diplostomum*.

Differential parasitism is present between sampled lochs in the Gairloch system. Discriminant analysis of the parasitism of brown trout in the lower-lying lochs (Lochen nam Breac, Loch Coire nah-Airigh, Loch Gharbe-Doire and Loch Feur) revealed increased number of cestode infection. Only trout in these lochs were infected with *Diphyllobothrium dendriticum, Eubothrium crassum, Eubothrium salvenius* and the acathocephalan species *Neoenchorynchus saginatus*. The discovery of cestode infection within these subpopulations was of increased importance because *D. dendriticum* is responsible for the food-borne tapeworm disease, Diphylobothriasis. The co-infection with other cestode species would point to an increased copepod presence within the low-lying lochs. The ingestion of infected copepods is a transmission route of cestode infection in brown trout. The presence of three-spine stickleback (*Gasterosteus aculeatus*) is also implicated as a direct factor in cestode infection within a freshwater body. A survey of possible stickleback presence may confirm this as a possible factor within these specific cestode infected lochs. Human intervention is another factor influencing host presence in the vicinity of cestode infected locales. The Gairloch recycling center is located near the low-lying lochs (latitude; 57.738481, longitude; -5.660867) provides a food source for scavenger birds such as the black-headed gull (*Chroicocephalus ridibundus*) and the common gull (*Larus canus*). Both species are definitive hosts for freshwater cestodes, in particular *D. dendriticum*. With the Gairloch system being highly oligotrophic, it is difficult to ascertain the tangible pathogenic effect of parasite infection. Behavioral feeding habits affects parasite infection communities in salmonids. Artic charr displays a richer parasite infective community in sympatric populations with brown trout, primarily due to the increased likelihood of zooplankton and copepod consumption (Knudsen *et al*., 2008). The presence of cestode infection due dietary adaptation of Gairloch subpopulations is in accordance with Kennedy *et al*. (1986) and Holmes (1990), who suggested that fish hosts with potential for the broadest dietary range would have an increased diversity of food transmitted parasites. The presence of the parasite *Diphyllobothrium dendriticum* and *Eubothrium crassum* in

particular indicates piscivorous feeding habits, with *D. dendriticum* having the ability to establish infection within prey fish compared to the primarily zooplankton based *D. ditreum* (Halvorsen and Wissler 1973; Curtis, 1981; Scholz, 2003).

The hill lochs within the Gairloch system offer variations in parasite-host interaction. The low lying-cestode infected lochs are smaller than larger elevated lochs. Smaller water bodies change the dynamic, which host and parasite can interact. Density of cercariae increases drastically within smaller volumes of water with little water movement creating a more viable transmission environment, in particular, *Diplostomum* eye flukes and other freshwater trematodes. Trout from larger lochs in the system had very little cestode infection, with the nematode *Eustronglyides* the only infecting parasite other than *Diplostomum baeri.* Nematodes in general display increased infection within larger fish populations compared to smaller populations. The nematode presence within these larger populations may reflect an increased number of larger sexually mature trout. Feeding habits of sexually mature trout are predominantly more aggressive than juvenile feeding on faster moving prey, increasing consumption potential of oligochaete worm species, an often-utilized paratenic host of *Eustronglyides* (Haugen *et al*., 2008 Oscoz *et al*., 2008). Factors associated with the exposure of trout hosts to parasites through feeding habitat and dietary behavior seem to be a very important structuring force of helminth communities in trout sub-populations.

Subpopulations of brown trout across the Gairloch system indicate a varied landscape of diversity in parasite infection. Primarily associated with host presence, different species exist within finite environments within the system. *Diplostomum baeri* shows an overwhelming presence across sub-populations of trout, accounts for the majority of raw parasite samples removed from Gairloch brown trout. Its high prevalence within freshwater fish populations has been seen worldwide, infecting not only brown trout but a multitude of freshwater fish species (Bakal *et al*., 2005; Larsen *et al*., 2005; Blasco-Costa *et al*., 2014b; Gendron and Marcogliese, 2017). Within the UK mainland only limited research has been performed to explain the high prevalence of this parasite in fish

populations (Stables and Chappell, 1986a; Whyte et al., 1991) exacerbated by the lack of molecular identification techniques to ascertain genetic diversity and true species within UK freshwater fish populations. With the molecular-based research being performed on European (Georgieva *et al*., 2013b) and north American (S. A. Locke et al., 2010a) fish populations, this is a prime opportunity to expand the knowledge of *Diplostomum* species richness and diversity within UK trout populations.

3. Diversity of Diplostomum. spp infecting *Salmo trutta* in Scottish Highland hill loch system.

3.1 Introduction

Early attempts to understand *Diplostomum* taxonomy produced conflicting results; Yamaguti (1958) placed the genus within the class Trematoda and order Digenea whilst Bykhovskaya-Pavlovska *et al* (1966) placed *Diplostomum* within the class Trematoda within the order Strigeidae. Because larval stages have simple morphologies and different life cycle stages have been the focus of separate taxonomic treatments (Valtonen and Gibson 1997) taxonomic placement on basis of morphological identification is difficult. Current methods of morphology-based identifications of eye flukes have attempted to utilize; site occupied, metacercarial morphology and infected host species as criteria to enable species characterization (Niewiadomska and Laskowski, 2002). More recently, approaches utilizing commonly-applied molecular markers have accelerated recognition of host-parasite associations and elucidated complex parasite life cycles of many freshwater digeneans (Blasco-Costa *et al*., 2014a; Georgieva *et al*., 2013a; T H Le *et al*., 2000; S. a. Locke *et al*., 2010; Moszczynska *et al*., 2009a)

The inability to identify larval infective stages has been a major impediment in assessing their role within wild fish populations (Sangster *et al.,* 2004) and evolutionary aspects of their host-parasite associations (Kalbe and Kurtz, 2006). Recent molecular studies within Europe and the Americas have provided significant evidence to characterize 24 species-level lineages of *Diplostomum* including three species complexes previously recognized as single species (Locke *et al*., 2010a) in Canada and twelve more in hosts from multiple life cycle stages in Europe (Georgieva *et al*., 2013b). Although limited in the geographic extent of the study area, these studies made use of use of nuclear markers (ITS1-5.8s-ITS2) to indicate a previously unrecognized diversity in mainland Europe, North America and within Arctic regions. The most recent comprehensive molecular analysis of 2000 newly acquired and published *Diplostomum* mitochondrial sequences taken from around the world indicated a total of 52 species (Locke *et al*., 2015). Only 33 of this 52 have

been identified in previous studies (Niewiadomska and Laskowski, 2002; Haarder *et al.,* 2013; Perez-del-Olmo *et al*., 2014). A population genetic study using both nuclear (ITS1-5.8s-ITS2) and mitochondrial genetic markers (*cox1, NADH, cytb etc.*) along with samples derived from a variety of geographical isolates would present a more complete phylogeography of the parasite and its true species richness worldwide. Moszczynska *et al.* (2009) used *Diplostomum* specific markers to build a comprehensive barcode library (58 linked ITS and 306 *cox*1 barcode sequences) of North American species of *Diplostomum*. Their study of the St. Lawrence River in Canada revealed high species diversity compared to previous morphology-based studies (Locke *et al*., 2010a, b). Along with the extensive research being performed within North America, more recent molecular, field=derived investigations have been based upon infections within Europe. Integrated ITS and *cox*1 sequence data from within fish intermediate hosts in Germany (Georgieva *et al*., 2013) and Spain (Perez-del-Olmo *et al.,* 2014) demonstrated genetic difference in nuclear marker sequences whilst still showing high morphological similarity, highlighting the need for molecular-based taxonomic studies.

Within the United Kingdom molecular identification of species of *Diplostomum* spp. has not taken place despite the presence of infection within wild fish (Wootten and Smith, 1980; McKeown and Irwin, 1997; Barber, 1997) and farmed salmonid populations (Mcguigan and Sommerville, 1985; Stables and Chappell, 1986; Whyte *et al.,* 1987). The increasing reliance on fish farm production, particularly within Scotland, makes the accurate and repeatable identification of one of the most ubiquitous parasites integral to understanding worldwide species richness and providing a potential tool for disease control methods.

3.2 Methodology

3.2.1 Sample collection

Molecular methods were used to characterize 60 parasite isolates dissected from eyeballs of whole *Salmo trutta* caught from 6 different lochs from the Gairloch region of North West Scotland (centered approximately 57º43’20.59”N, 5º38’6.85”W) and donated by

anglers (2013-2016 inclusive). Dissection and removal of all parasites were in accordance with Environment Agency fish health check protocol (Environment Agency, 2008). Extracted parasites were stored in 70% ethanol. Each collection site was represented by 10 metacercariae isolates from 10 different infected fish. These site were denoted as Loch na Feithe Mugaig (North; 745, East; 856), Loch Airigh a' Phuill (North; 846, East; 755), Loch Doire na h-Airigh (North; 741, East; 874), Lochan nam Breac (North; 785, East; 813) and Loch na h-Oidche (North; 668, East; 884).

3.2.2 PCR amplification

Total genomic DNA was extracted using protocol of Qiagen^{TM} Blood and Tissue extraction kit. For the amplification 410bp of the cytochrome c oxidase gene (*cox1*) and 780bp of Internal transcriber spacer (ITS1-ITS2) region primers were designed using primer-BLAST software from NCBI (www.ncbi.nlm.nih.gov/tools/primer-blast/). Partial *cox*1 fragments were amplified using the diplostomid-specific PCR primers (Moszczynska *et al.*, 2009): mplatCOX1df (forward; 5'-TGT AAA ACG ACG GCC AGT-3') and MplatCOX1dR (reverse; 5'-CAG GAA ACA GCT ATG ACT-3'). In addition to primers used for PCR amplification, specific primers were designed to aid in specificity and accuracy in association with further sequencing technologies, these were MplatCOX1SEQdF (5'-TGTAAAACGACGGCCAGTTTWCITTRGATCATAAG-3') and MplatCOXSEQdR (5'-CAGGAAACAGCTATGACTGAAAYAAYAIIGGATCICCACC-3'). The ITS1-5.8S-ITS2 cluster was amplified for a subset of isolates from each *cox*1 derived lineage using the primers of Galazzo *et al*. (2002) MplatdiploITSF (forward; 5'-AGG AAT TCC TGG TAA GTG CAA G-3') and MplatdiploITSR (reverse; 5'-CGT TAC TGA GGG AAT CCT GGT-3').

PCR reactions were performed for each gene fragment using 12.5 µl hot start taq polymerase; Thermo–Start^{R}PCR master mix (0.625 Units of Taq DNA polymerase, 1X reaction buffer, 0.2 mM of each dNTP and 1.5 mM $MgCl_2$) and 1–2 ng/µl of DNA. The template PCR program used for all isolates was: 95° for 15 minutes; 40 cycles at 95°C, 1 min; 55-58°C; 72°C, 1 min; 72°C for 7 minutes. Annealing temperatures were optimized on the primer set with platydiploITS using 57°C and platydiplocox1 52°C. Final reactions

were made up to 25 µl with PCR-grade water. Reactions were performed using a Techne Prime 96 well thermal cycler (Techne™) and 5 µl of each amplicon was visualized in 1 % agarose gels stained with SafeView nucleic acid stain (nbsbiologicals™) under UV using Licorgel documentation system. The remaining 20 µl PCR products were sequenced using Sanger sequencing protocol at the DNA sequencing facility of the Natural History Museum, London, using the PCR primers with Fluorescent Dye Terminator Sequencing Kits (Applied Biosystems™); sequencing reactions were run on an Applied Biosystems™ 3730XL automated sequencer.

3.2.3 Alignments and Data analysis

All sequences analyzed in the dataset were sampled from the intermediate metacercarial stage of the life cycle. All sequences from the Gairloch data set represent *Diplostomum* from a single infected individual, the dataset including individuals from 6 different loch subpopulations with 10 samples in each loch.

Both published and the newly generated *cox1* and ITS sequences were aligned using Clustal W implemented in BioEdit software. Both *cox1* and ITS alignments were examined using 60 newly-acquired sequences derived from Gairloch isolates and previously published sequences retrieved from Genbank (https://www.ncbi.nlm.nih.gov/genbank/) from European and north American isolates of *D. baeri, D. spathaceum, D. pseudospathaceum, D. mergi* and an outgroup of *Tylodelphys clavata*.

To deal with heterozygous sites arising from direct sanger sequencing PHASE analysis package as part of DNAsp software (Librado and Rozas., 2009) was utilized to generate a dataset arising from heterozygote alternative sequences and initial sequence. Prior to phylogenetic tree construction, partitioning of sequence within BEAST software package (Bouckaert *et al*., 2014) ignores ambiguous sites within heterozygous sequences. The removal of heterozygous sites within a sequence will result in the data set with less homoplasy. However entering heterozygous sites, as part of the alignment will affect branch support for each allele in a clade, with artificial site diversity lacking autapomorphies within true sequence.

Species boundaries were inferred using via Maximum likelihood (ML) method using Bayesian inference criterion (BIC) in Beast 2.0 software (Bouckaert *et al*., 2014) using general time reversible models including estimates of site-rate heterogeneity (GTR+G) for the *cox*1 dataset and the Kimura-2 parameter model including estimates among-site rate heterogeneity and invariant site (K2P+I+G). Heuristic stop criteria for ML searches were fulfilled after 120 (*cox1* dataset; best tree 1,234,43) and 110 (ITS1; best ML tree 1,493,54). Additional saturation analysis was performed on the *cox1* phylogeny. The accuracy of phylogenetic reconstruction depends on the correct identification of homologous sites by sequence alignment and the presence of substitution saturation. The divergence that is present in a sequence can be neither too conserved nor too diverged to experience substitution saturation. Saturation decreases the phylogenetic information contained in sequences leading to problems in the trustworthiness in phylogeny. DAMBE (Xia, 2001) is a GUI-Windows based program used to measure the frequency of transitions and transversions across a sequence alignment against corrected genetic distance. This gives confirmation of accuracy of mitochondrial trees in particular, where the rapid accumulation of mutations across mitochondrial genes can lead to saturation.

Diversity data were acquired through the use of DNAsp software package (Librado and Rozas, 2009), developed for comprehensive analysis of DNA polymorphism data. The genetic diveristy profile is defined in 5 parameters;

- **S**; segregating sites, this values denotes the amount of ambiguities that exist in a pairwise comparison between the sequences and is taken into account in construction of division of haplotypes, in network construction
- **h;** Haplotype number, denotes the number of haplotypes that exist in the dataset, this refers to the SNP mutations that are shared within a select group of sequences.
- **K;** this denotes the average number of nucleotide differences that occurs in a pairwise comparison of alignments within a dataset (Tajima, 1983). This is computed in response to stochastic variance (Tajima, 1993 equation 14), Sampling variance (Tajima, 1993 equation 15) and Total variance (Tajima 1993 equation 13) for no recombination parameters. Stochastic variance (Tajima, 1993 equation 17), Sampling variance (Tajima, 1993 equation 18) and Total variance

(Tajima, 1993 equation 16) for free recombination parameters, hence K is denoted as 'average' number of nucleotide differences.

- **Pi;** Nucleotide diversity, the average number of nucleotide substitutions per site in a pairwise comparison across the dataset (Nei, 1987). Unlike **K**, **Pi**, is adjusted through the use of Jukes-Cantor (1969) correction. The correction is performed in a stepwise manner in each pairwise comparison and **Pi** estimate values are obtained as an average of the values in all compared sites.
- **hd;** Haplotype diversity is a measure of the uniqueness of a particular haplotype within a given population sample. It can be described as **Pi** with pairwise comparison existing between haplotypes rather than individual sequences. The haplotype diversity is computed as:

$$H = \frac{N}{N-1}(1 - \sum_i x_i^2)$$

Where x is the relative haplotype frequency existing of haplotype grouping in the sample and N is the sample size.

Haplotype network analysis was performed using PopArt Haplotype network (Leigh and Bryant, 2015), distance was computed using minimum spanning network analysis to provide iteration data between haplotypes. A minimum spanning network is a contructed network where the total weight of variance in the edge points is as small as possible. This increases the likelihood of network joining particularly for clonal organisms. The principal exists in a distance based method, first a matrix network of pairwise differences among haplotypes is built and then the shortest paths that link the observed haplotypes is found where the length of connection is determinant on the number of pairwise differences (Kurksal, 1956).

3.3 Results

3.3.1 Phylogenetic analysis

ML analyses based on the *cox1* indicated all eye flukes from Gairloch system brown trout belonged to the *Diplostomum baeri* multi-lineage species complex. Analysis of the phylogenetic relationship of Scottish-derived samples and European-derived sequences reveal that *Diplostomum baeri* exhibits a multi-lineage species complex within Scotland. According to both nuclear (Fig 2) and mitochondrial (Fig 1) genetic markers, samples collected from the Gairloch hill loch system reside within the same clade as previously-published sequences of *D. baeri*. Scottish isolates were placed alongside isolates of *D. baeri* from Icelandic and German freshwater systems.

The first *D. baeri* lineage described was made up entirely of Canadian-derived samples representing samples infecting intermediate host species *Sander vitreus, Salvelinus alpinus* and *Salmo trutta*, thus representing the North American clade of *D. baeri* (Fig 1). Samples deriving from Lineage 2 represented the first clade including samples from the investigation in Gairloch system. This comprised entirely salmonid-derived samples including Scottish and Icelandic *Salmo trutta* and Icelandic *Salvenius alpinus*. The make up of Lineage 3 is entirely made up of metacercariae infecting *Perca fluviatilus* within Lake Constance, Germany. Lineage 4 is made up of a diverse mix of isolates from mixed localities including Icelandic, German and Scottish locations, within this clade all the metacercariae have derived from *S.trutta* host species.

The phylogenetic hypothesis inferred through the ITS1-5.8s-ITS2 dataset differed from the mitochondrial phylogenetic analysis (Fig 2). The phylogenetic reconstruction indicated a collapse of lineages displayed in the mitochondrial tree. The use of nuclear markers show diminished diversity compared to mitochondrial markers. Comparisons of the nuclear markers to previously sequenced European samples were few with the majority of published sequences being from mitochondrial markers (ITS-5.8s-ITS2: 22 and *cox1*: 121 (https://www.ncbi.nlm.nih.gov/nuccore). Phylogenetic relationships still placed the majority of the Gairloch derived samples within the *D. baeri* barring samples LAP 51, LNB 03 and LNB 16, however BLAST during construction of alignments these samples shared 98% similarity with both previously published *D. baeri* and *D. pseudospathaceum* sequences.

In comparison to the mitochondrial tree, species specificity does not play a role within defined clades. Lineage 3 and 4 in *COI* ML tree indicated species specific *Salmo trutta and Perca fluviatilus* clades, ITS markers placed both *Perca fluviatilus* and *Onchorynkus mykiss* infecting metacercariae with *Salmo trutta* infecting *D. baeri*. Published data from mainland Europe indicated a majority shared ITS haplotype, Scottish samples were unique haplotypes with no shared haplotypes with mainland continental European samples.

Phylogenetic construction using *cox1* genetic marker was corroborated using Ts/Tv plot to test for mutational saturation within the marker. The plot has tested the presence of multiple substitutions occurring at the same degenerate codon position, which was tested on 3rd codon. The data presented here does show negligible saturation of transitions with the line not demonstrating a strict straight line. The result of potential saturation existing within the third codon of the *cox1* dataset could negate the use of 3rd codon within deeper phylogenies build using the *D. baeri cox1* gene (Figure 10).

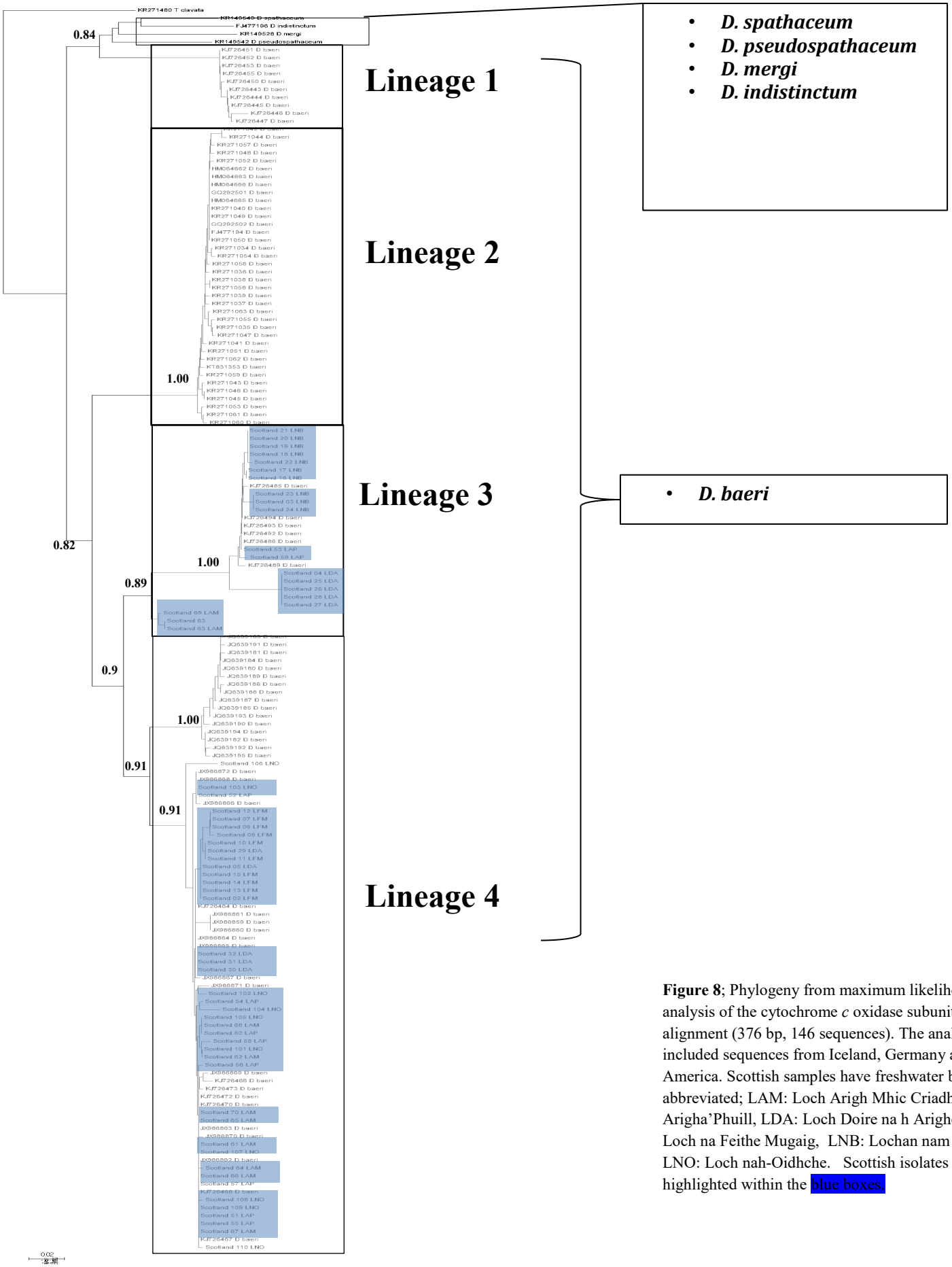


Figure 8; Phylogeny from maximum likelihood derived analysis of the cytochrome *c* oxidase subunit 1 sequence alignment (376 bp, 146 sequences). The analysis included sequences from Iceland, Germany and North America. Scottish samples have freshwater bodies abbreviated; LAM: Loch Arigh Mhic Criadh, LAP: Loch Arigha'Phuill, LDA: Loch Doire na h Arighe, LFM: Loch na Feithe Mugaig, LNB: Lochan nam Breac, LNO: Loch nah-Oidhche. Scottish isolates are highlighted within the blue boxes.

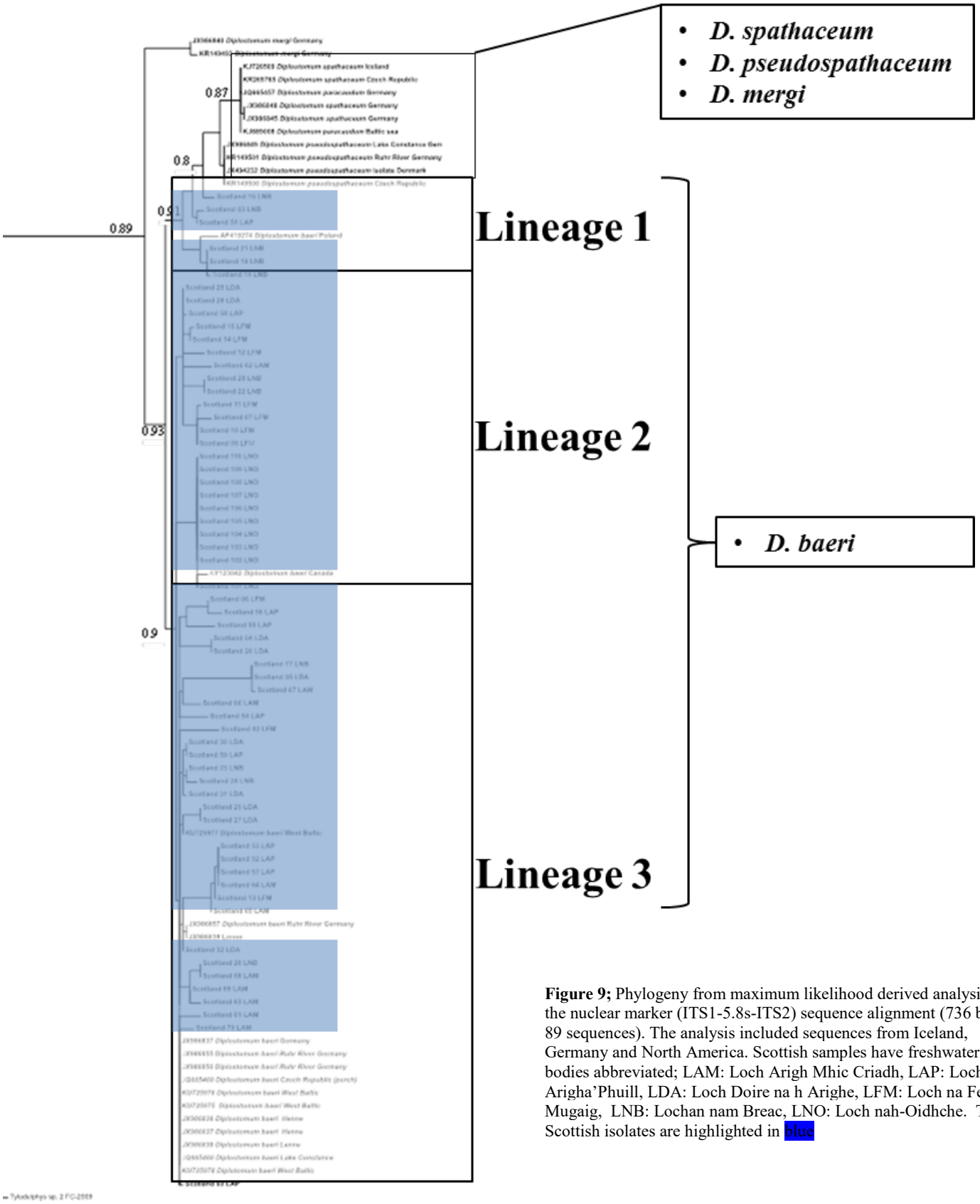


Figure 9; Phylogeny from maximum likelihood derived analysis of the nuclear marker (ITS1-5.8s-ITS2) sequence alignment (736 bp, 89 sequences). The analysis included sequences from Iceland, Germany and North America. Scottish samples have freshwater bodies abbreviated; LAM: Loch Arigh Mhic Criadh, LAP: Loch Arigha'Phuill, LDA: Loch Doire na h Arighe, LFM: Loch na Feithe Mugaig, LNB: Lochan nam Breac, LNO: Loch nah-Oidhche. The Scottish isolates are highlighted in blue

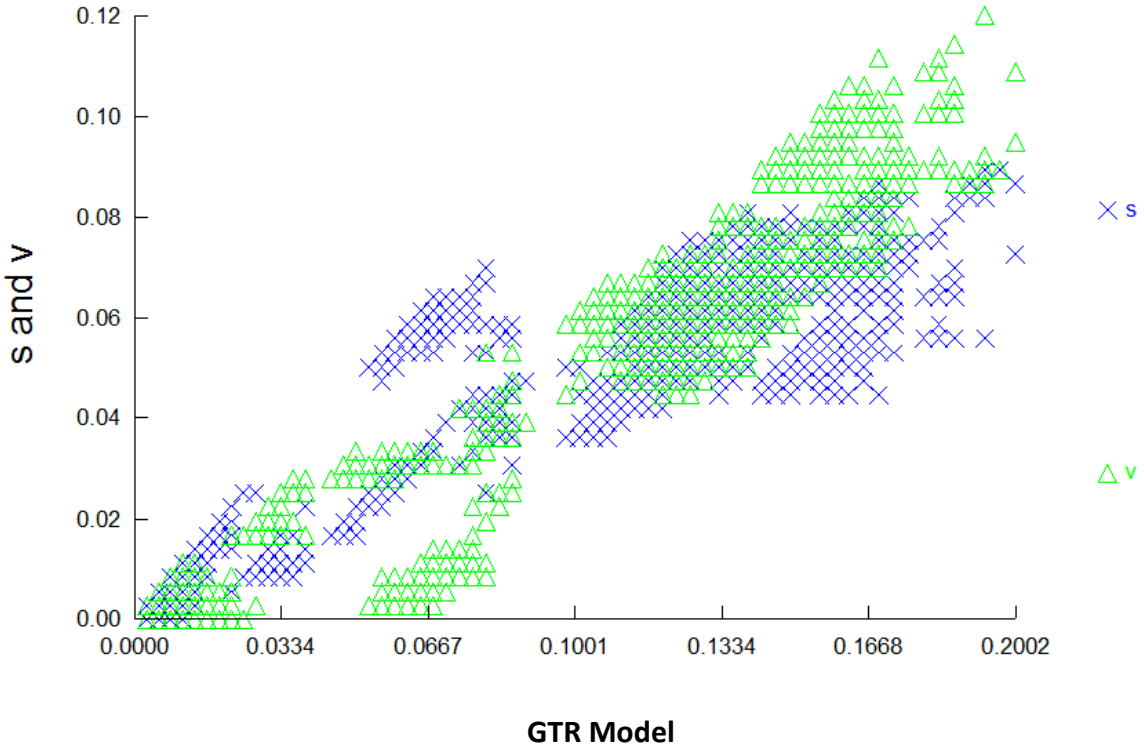


Figure 10: DAMBE plot for saturation test. X axis represents genetic distance according to phylogeny used in tree reconstruction. Y axis represents Ts/Tv number as genetic distance increases, tested across the 3rd codon within *cox1* alignment.

3.3.2 Diversity indices of *cox1* and ITS and haplotype network analysis

Haplotype network construction of *cox1* haplotypes (Table 5 and Figure 8) showed complex inter-specific relationships between lineages across mainland continental European and North American derived isolates. Segregation site number between haplotypes was low across the entire lineage data set compared to geographic lineage genetic diversity (table 3). Salmonid-infecting lineages possessed increased numbers of unique haplotypes compared to perch-infecting lineage (3), lineage 2 (27) and lineage 4 (45) had the highest number of haplotypes compared to all other lineages, however the increased number of hapylotypes was not mirrored in the genetic diversity existing between haplotypes with haplotype diversity being the lowest (0.489). Similar patterns appeared within the ITS dataset (Table 6 and Figure 8) with low diversity being demonstrated across the entire dataset. In particular lineage 3 where the largest amount of haplotypes, however the large amounts of haplotypes do not show an increased level of intra-clade haplotype diversity. The increased number of haplotypes within the Gairloch system compared to the other locality isolates may not be representative in terms of nuclear genetic diversity. With an overall lack of published nuclear sequences in other papers the increased diversity of Gairloch isolates is most likely due to the number of samples used compared to other localities.

The genetic diversity profile generated from *Diplostomum baeri* sequence dataset demonstrated genetic diversity between different sampling sites within the study area. The *cox1* dataset shows the location with the highest number of individual haplotype and the greatest haplotype diversity being Loch na h-Oidche outflow (LNO) (H; 9, HD; 0.978) and Loch Arigh a'Phuil (LAP) (H;8 and HD 0.963) this is mirrored in the average amount of nucleotide diversity per site, with both sites having the highest K value in the dataset (LNO; 14.486 and LAP; 14.822). The lowest amount of haplotype diversity is Loch na Feithe Mugaig (LFM) (H;4 and HD; 0.711), this is the same for the lowest overall diversity in the location (Pi; 0.00414).

Within the ITS dataset the two locations which were the most genetically diverse were the same locations indicated using mitochondrial markers; LAP and LNO indicating the

highest number of individual haplotypes (LAP; 9 and LNO; 9), haplotype diversity (LAP; 0.955 and LNO; 0.988) and high diversity per site (LAP; 20.067 and LNO; 20.003). The sites also display the highest overall nucleotide diversity (LAP; 0.0209 and LNO; 0.031). The location that displays the lowest amount of nucleotide diversity was Loch na Feithe Mugaig (LFM), which had the lowest haplotype number (6), lowest haplotype diversity (0.907) along with the lowest K value (15.333) and Pi value (0.01589).

ITS	S	h	K	Pi	hd
Lineage 1	24	6	10.14286	0.00912	0.988
Lineage 2	24	10	6.77856	0.05607	0.567
Lineage 3	87	21	15.99878	0.18999	0.493
cox1					
Lineage 1	6	5	1.55556	0.00515	0.511
Lineage 2	21	27	6.15508	0.04561	0.618
Lineage 3	9	11	5.73331	0.07789	0.412
Lineage 4	87	45	15.9987	0.08642	0.489

Table 4; Displaying evolutionary properties between populations within Gairloch *Diplostomum* dataset using DNA barcoding markers (*cox1*) and nuclear marker (ITS1-5.8s-ITS2)**.S** - Number of segregating sites, **Pi** – nucleotide diversity, **K** – average number of nucleotide differences per site, **h** – number of haplotypes, **hd**– diversity between haplotypes observed. Lineage 1-3 of *cox1* and 1-3 in ITS are in correlation with the lineage described and highlighted within respective phylogenies.

ITS						
Site	***n***	**S**	**h**	**K**	**Pi**	**hd**
LAM	10	24	6	8.089	0.02186	0.844
LAP	10	41	8	14.822	0.04006	0.963
LDA	10	3	4	1.533	0.00414	0.778
LFM	10	44	4	4.222	0.06541	0.711
LNB	10	4	4	1.956	0.00529	0.773
LNO	10	28	9	14.486	0.01213	0.978
cox1						
LAM	10	63	6	16.5	0.01728	0.933
LAP	10	79	9	20.067	0.0209	0.955
LDA	10	63	8	16.822	0.01743	0.956
LFM	10	57	6	15.333	0.01589	0.907
LNB	10	60	9	17.267	0.01795	0.977
LNO	10	70	9	20.003	0.031	0.988

Table 5; Displaying evolutionary properties between populations within Gairloch using nuclear markers (ITS1-5.8s-ITS2) and mitochondrial markers (cytochrome oxidase 1).S - Number of segregating sites, Pi – nucleotide diversity, K – average number of nucleotide diversity per site, h – number of haplotypes, hd– diversity between haplotypes observed.

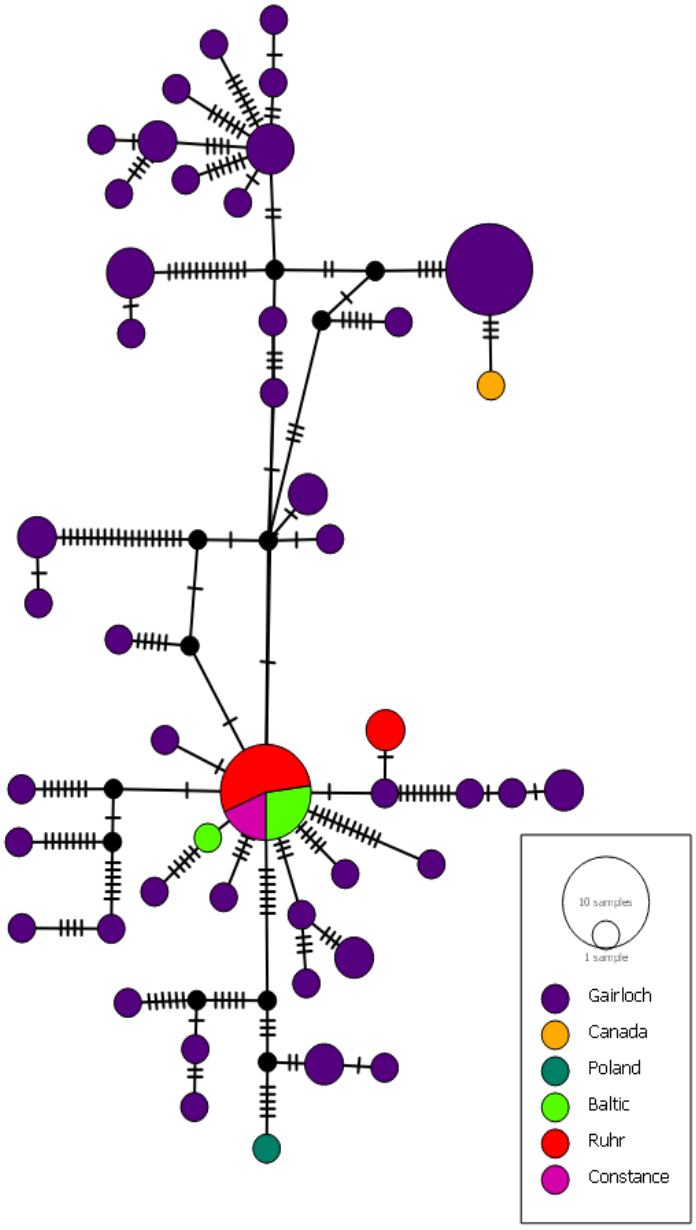


Figure 11; Haplotype network analysis of ITS1-5.8s-ITS2 for 77 isolates using TCS spanning network. Isolates are divided using geographical locations of isolates from North America, Poland, West Baltic, Ruhr River, Lake Constance and Gairloch. Segregating sites are denoted via hatch marks between haplotypes, the number of individuals with shared haplotype are indicated via size of the individual pie charts.

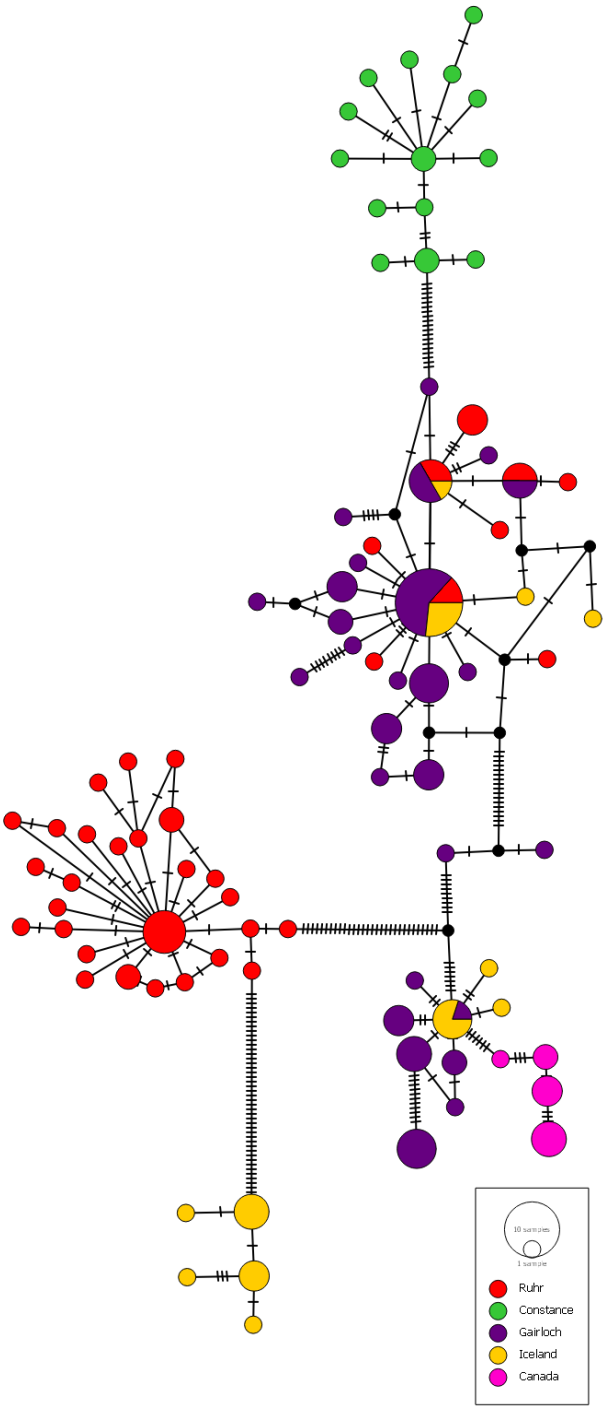


Figure 12; Haplotype network analysis of cytochrome c oxidase subunit 1 for 146 isolates us TCS spanning network. Individual haplotype charts represent geographical isolates from Iceland, River Ruhr, Lake Constance, and Gairloch. Segregating sites are denoted via hatch marks upon adjoining lines between haplotypes, the number of individuals with shared haplotype are indicated via size of individual pie charts.

3.4 Discussion

This study is the first to provide estimates of species diversity of *Diplostomum* utilizing molecular methods within a United Kingdom fresh-water ecosystem. It is also the first analysis of purely Brown trout-infecting metacercariae and the phylogenetic analysis and genetic diversity within a hill loch environment. Within the hill loch system, diversity would be directly correlated by the system's greater or lesser ability to support specific intermediate host species, presenting varying levels of life-cycle bottleneck in relation to the transmission of certain species of *Diplostomum* directly influence levels of infection in intermediate fish host.

Gairloch presents a unique natural setting to study the effects of geographic isolation of populations on digenean diversity. The northwest Highlands of Scotland are isolated uplands with a history of extreme human intervention in landscape management culminating in almost complete deforestation with consequent removal of nutrients from the loch system and surrounding area. Previous studies carried out in elevated latitude environments indicated an underestimation of species richness and overall genetic diversity of infective eye fluke species (Kristmundsson & Richter, 2009; Blasco-Costa *et al*., 2014). Compared to Icelandic lake populations (Blasco-Costa *et al*., 2014a; Kristmundsson and Richter, 2009), Scottish brown trout revealed significantly lower species richness of *Diplostomum spp*. Elsewhere within Palearctic freshwater systems brown trout hosts three species of *Diplostomum* whilst central European populations host two. The Gairloch study area had more sampling sites (14) than cited Icelandic and German studies, with six freshwater sites revealing a lower level of species diversity compared to the Icelandic study (4 sites) (Blasco-Costa *et al*., 2014a) and German Study (2) (Georgieva et al., 2013a; Selbach et al., 2015a).

Although taxonomic identification of metacercariae indicated lower species diversity than previous studies within brown trout the *D. baeri* found within the Gairloch study area did present as a "multi-lineage species complex" with multiple lineages of the parasite within Gairloch. (fig 7 & 8). The term 'multi-lineage species complex' is used in taxonomy as a way of describing phylogenetic placement in 3 discerning statements; the group of species may represent more than one species (Mills *et al.,* 2017)*,* this cannot be discern by morphological or molecular means (Blasco-Costa *et al.,* 2014) and that

grouped species within the complex are related in some way (Fontaine *et al*., 2015) Within the world of parasitology the term species complex is utilized to explain taxonomic topology resulting from numerous hypothetical assumptions e.g., the high probability of genomic introgression (Webster et al., 2013; Fontaine et al., 2015) , recent rapid radiation (Bush et al., 2018) and definitive host movement (van Paridon et al., 2017; Rosenkranz et al., 2018). With respect to *Diplostomum* both *D. baeri* and *D. mergi* have been described as presenting a multi-lineage species complex topology (Blasco-Costa *et al*., 2014a; Georgieva *et al*., 2013a; Locke *et al*., 2010; Locke *et al*., 2013). *Diplostomum baeri* was first described in Europe demonstrating species-specific 'trout' and 'perch' lineages (Behrmann-Godel, 2013). Species specificity has been a driver in speciation between other species of freshwater trematodes (Trieu *et al*., 2015; Pinto *et al.*, 2016; Pinto *et al.,* 2018). Species specificity may also be a speciation driver within *D. baeri* where host specific lineage may become increasingly reproductively isolated leading to eventual speciation. To examine the delineation between lineages that currently exists in within *Diplostomum* phylogeny marker choice could provide a useful tool. Nuclear markers used within this study do show significantly less diversity between samples compared to mitochondrial markers. Although it does show increased diversity compared to other localities, this may be due to the number of samples used in the Gairloch study, with a scarcity of ITS nuclear samples found in other studies. The use of more diverse mitochondrial markers may be better utilized to investigation multi-lineage relationships. Lactate dehydrogenase (nad1, nad4 and nad5) have been used to elucidate taxonomic relationship within freshwater trematode species *Orientobilharzia turkestanicum* (Li et al., 2008), *Clonorchis* (Park, 2007) and *Fasciola* (Hu-qui *et al.,* 2009). Lactate dehydrogenase markers present as some of the most divergent markers within mitochondrial genomes of trematodes (Le et al., 2002; Shekhovtsov et al., 2010; Liu et al., 2014; Brabec et al., 2015a). Although the large number of mutations that would accumulate within lactate dehydrogenase genes may prove problematic when using the marker on deeper phylogenies, the rapidity of their accumulations of mutations may make these markers useful for the delineation of closely related taxa.

The current state of molecular taxonomy indicates that the *D. baeri* multi-lineage species complex is divided into eight molecularly-characterized species lineages, five recorded within North America, two lineages shared between Europe and Iceland and one in Iceland only. Phylogenetic analysis of the *D. baeri* complex within Gairloch show that unlike in Iceland and North America there is no true 'Scottish lineage' of *D. baeri* that resides within Scotland. The panmitic admixture of haplotypes infecting the Gairloch populations of brown trout could in part be due to its placements within the Atlantic migratory bird pathway (Elphick, 2007). Transitory bird populations of the piscivorous *Gavia arctica* (black-throated diver) and *Gavia stellata* (red-throated diver) inhabit the northwest highlands during summer months, with around 17,000 *G. stellata* wintering in the UK and 1,600 permanent mating pairs. The relatively rare *G. arctica* has 560 pairs of UK wintering birds and 250 permanent mating pairs residential in the UK (O'Brien *et al*., 2008). UK populations of both birds are entirely located within the Northwest Highlands of Scotland and Northern Ireland, and utilize elevated hill lochs to feed and reproduce (Mudge *et al*., 1998). The migration of piscivorous birds provides a major route for the spread of *Diplostomum* in a country- and continent-wide fashion. With the parasites' adult stage residing in the definitive bird host, infected birds will disseminate eggs into freshwater locales. In the case of migratory birds, infection can be spread across great distances with parasites being trans-placed between wintering and breeding grounds.

Infective dynamics of *Diplostomum* within Gairloch represents similar characteristics associated with genetic diversity between water bodies. The sampling site with the greatest number of haplotypes and diversity between the haplotypes was LNO (Loch na h-Oidche), the site represented the only one which was a riverine environment. Brown trout within riverine environments have increased immune gene heterogeneity compared to lacustrine populations, probably because of a wider variety of pathogens being encountered within riverine environments (Fraser *et al*.,2011). This is in part due to riverine environments providing a more diverse habitat for varied intermediate snail host species, which would, in turn, suggest the likelihood of more genetically diverse cercaerial infective stages being present. Also riverine populations are more diverse population genetically than lacustrine (Hindar *et al*., 1991). *Diplostomum baeri* shows a

high infective success, with the ability to utilize a wide range of freshwater fish species as intermediate host, part of its ability to infect such a wide range requires *it* to successfully evade a wide variety of individual host immune systems, selectively increasing the diversity of successfully-infective haplotypes of *D. baeri*. When compared with European isolates in haplotype network analysis (fig 10) hill loch mitochondrial markers show a diversity of haplotypes across all sampled areas with isolated haplotype clusters associated with host species infected and not geographic isolation. All salmonid-infecting species had shared haplotypes across all geographic isolations.

3.5 Concluding remarks

In conclusion, findings within Scottish Highland hill loch system suggest a low level of biodiversity of *Diplostomum* infecting brown trout hosts inhabiting the Gairloch hill lochs. The low level of species richness of *Diplostomum* could in part be due to the highly oligotrophic nature of the loch system acting as a life cycle bottleneck upon intermediate gastropod and salmonid host species. *Diplostomum baeri* represents the most prevalent parasite within the loch system and is the most widespread species of *Diplostomum* within salmonids worldwide. The presence of multiple lineages within the species complex suggests its infective success is, at least in part, related to its within-taxon diversity. Although the species shows high level of speciation, definitive host movement is probably sufficient to ensure that true reproduction isolation of geographic haplotypes and thus full speciation of lineages is unlikely to occur.

D. baeri was the most prevalent parasite infecting brown trout within the Gairloch hill loch system. In-depth genetic analysis allowed accurate parasite identification and diversity analysis. Use of both mitochondrial and nuclear markers confirmed *Diplostomum baeri* species identification. Mitochondrial DNA has proved to be suitable for barcoding parasite species in highly informative phylogenetic and phylogeographic studies (Hendrich et al., 2010; S. A. Locke et al., 2010a; Kudlai et al., 2017b). Nevertheless using both nuclear DNA and mitochondrial DNA improves the power of phylogenetic and phylogeographic tests significantly and highlights the limitations of

studies relying purely on mt DNA markers. A phenomenon known as mito-nuclear discordance (Morales et al., 2015; Patten et al., 2015; Payseur and Rieseberg, 2016; Perea et al., 2016) (Toews and Beresford, 2012) resulting in; incomplete lineage sorting (Lawton et al., 2017), adaptive introgression of mtDNA (Ziętara et al., 2007), hybrid-zone geography (Moné et al., 2015), frequent male population crash (Jiggins, 2002) and parasite host rapid radiation (Webster et al., 2013). Multiple gene analyses are strongly recommended for both phylogenetic and phylogeographic studies of parasites in general (Wiemers and Fiedler, 2007; Kodandaramaiah et al., 2013), to avoid such pitfalls, particularly when studying such a prevalent parasite as *Diplosotomum*.

The impact of *Diplostomum* infection within the wild environment is difficult to ascertain, however on farmed trout populations epidemics may become problematic. With infected fish becoming less likely to feed, overall farming yield will eventually diminish. The farming of salmonids has proven lucrative in recent years making the understanding of parasites infecting oft-farmed species of salmonids crucial to continued farming success and for the many isolated communities relying on them. The identification of *Diplostomum* species can not only expand the understanding of the species diversity and richness worldwide, but also provide methodology to accurate identify salmonid parasites. The identification of freshwater parasites can aid in future monitoring and mapping of parasites throughout a system not just for *Diplostomum* but also for the varied parasitic fauna that infect residential and anadromous trout.

4. Molecular characterization of *Diphyllobothrium dendriticum* infection in UK populations of Salmo trutta

4.1 Introduction

Human diphyllobothrosis is a fish-borne zoonotic disease distributed worldwide caused by cestodes belonging to the genus *Diphyllobothrium* (Scholz *et al*., 2009). The life cycle of *Diphyllobothrium* involves two intermediate hosts including planktonic copepods and freshwater fish (Scholz and Kuchta,. 2016). Second intermediate host species include; perch (*Perca* spp.) (L. 1758), salmonids (*Salmo* spp.) (Cuvier, 1816), pike (*Esox lucis*) (L. 1758) and pikeperch (*Sander lucioperca*) (L. 1758) (Anderson *et al*., 1987), within which the parasite matures into plerocercoid stage of the lifecycle. Definitive hosts, such as carnivores, cetaceans, birds and humans become infected after ingesting the fish intermediate host. Adult worms attach to the intestinal wall causing abdominal pain and diarrhoea with chronic infections leading to B12 deficiency and pernicious anaemia (Von Bonsdorff, 1958).

Areas exhibiting high prevalence of diphyllobothriasis occur predominantly within cold climates with human pathogenesis distribution closely linked to cultural eating habits of populations (Scholz *et al*., 2009) The increasing mainstream popularity of sushi, sashimi, carpaccio and ceviche increases the risk of infection considerably. Concomitant with increased migration of food culture within the western world (Scholz and Kuchta, 2016) *Diphyllobothrium latum* and *Diphyllobothrium dendriticum* represent the most prevalent human infective fish tapeworm within northern Europe. The original distribution of *D. dendriticum* worldwide human infection is circumboreal, with a further spread of infection into South America (Santos and Faro, 2005; Mercado *et al*., 2010). *D. dendriticum* tends to predominate within Scandinavia (Kuhn *et al*., 2016; Borgstrøm *et al*., 2017) and alpine European regions (Dezfuli *et al*., 2012.) utilizing primarily salmonids and coregonids (Bylund and Curtis 1991) and as intermediate fish hosts. North America represents another endemic area, with infections found in Inuit populations of

Alaska and Canada (Mudry and Anderson, 1977). Despite an increase in food health regulations across the European Union, (Murrell and Cross, 1997) re-emergence of infection has occurred, with reports of increased infection prevalence within the alpine region of Italy, France and Switzerland (Peduzzi and Boucher-Rodoni, 2001; Yera *et al*., 2006) where surveys of human infection numbered 200 reported cases within the time period of 1987 to 2002 and 330 between 2002 and 2007 (Scholz, 2009; Wicht, 2009)

A total of 14 out of 50 species of *Diphyllobothrium* are pathogenic to humans (Kamo, 1978; Scholz *et al*., 2009). Accurate identification of parasite species is a pre-requisite to understanding disease epidemiology and focus potential control measures, especially when identifying parasites which have a wide intermediate host range and a medical risk (Guo *et al*., 2012). Most morphological differentiation is based around the shape and variability of the scolex of the adult worm, which has been shown to vary not only between but also within species, (Anderson *et al*., 1987; Mercado *et al*., 2010), making accurate species identification problematic. Molecular-based taxonomic identification has become essential to identify cestode species (Yamasaki and Kuramochi, 2009; Wicht *et al*., 2010; Thanchomnang *et al*., 2016). Molecular techniques have also been used to help understand the re-emergence of *Diphyllobothriasis* and newly acquired species infections within previously non-infected regions (Peduzzi and Boucher-Rodoni, 2001; Wicht *et al*., 2008, 2010; Fang *et al*., 2015).

In an attempt to resolve taxonomic knowledge of *Diphyllobothrium* and improve the accuracy of diagnostic identification procedures a number of molecular markers have been utilized. Markers include the use internal transcriber spacer regions ITS1 and ITS2 (Wicht *et* al., 2010; Pastor-Valle *et al*., 2014;) the lactate dehydrogenase gene NADH (Yera *et al*., 2006, 2008) and cytochrome oxidase cox1 (Yera *et al*., 2006; Guo *et al*., 2012; Thanchomnang *et al*., 2016). Of these, the mitochondrial marker cox1 has been primarily used in molecular diagnostic identification of adult worm infections around the world. Nuclear markers ITS1 and ITS2 have been used primarily to identify the intermediate infective plerocercoid stage within fish. Phylogenetic relationships within *Diphyllobothrium* are not fully understood due to most published sequences derive from medically important species. These have however demonstrated that there exists a close

relationship between *Diphyllobothrium dendriticum, Diphyllobothrium nihonkaiense, Diphyllobothrium latum* and *Diphyllobothrium ursi* (Kim *et al*., 2007; Year *et al*., 2008) Within the United Kingdom the first reported *Diphyllobothrium* infection in fish was in Loch Morar, Inverness-shire (Chaloner, 1912). Since then large-scale intermediate host infections have been seen in South Wales (Duguid and Sheppard, 1944) where large numbers of infected fry died. Although only a few parasite surveys have taken place in the UK, morphological identification of the plerocercoid stage of *D. ditreum and D. dendriticum* has been isolated within UK Rainbow, Brown and Sea trout (Thomas, 1964; Urquart *et al*., 2010). Both species were also identified using molecular techniques in a small sample of varied freshwater fish species in Scottish Lochs (Scholz, 2009).

Despite the relative importance of *D. dendriticum* little is known about its epidemiology and distribution across Europe. Previous studies of parasite movement throughout Europe have isolated ungulate migration as a key factor in the eastward movement of *Orientobilharzia turkestanicum*, now maintaining infective levels in Hungarian populations (Lawton and Majoros, 2013). The eastward movement of ungulates has also been proven to increase the overall parasite diveristy that existes within ungulates throughout Eurasia (Teitelbaum *et al*., 2018). The ancient east-ward migratory movement of animals throughout Europe could also play a role in *D. dendriticum* transplacement within the continent. Other species of *Diphylobothriid* have isolated carnivourous mammals as crucial hosts in maintatining the parasite life cycle within habitats in Canada (Gau *et al.,* 1999; Yamasaki *et al*, 2012) and eastern Russia (Arizono *et al.,* 2008). The eastern movement of *Ursus arcos* (Gray, 1867) (brown bear) has already been confirmed through dating of mitochondrial genome sequences as an eastward moving mammal with Estonian and Finnish populations deriving from Russian populations (Saarma *et al*., 2007). The ancient eastern movement of extinct *Ursus spelaeus* (Rosenmuller, 1794) (Cave bear) was also proven through the molecular dating of mitochondrial sequences of eastern most UK and Anodran populations deriving from populations in the Carpathian and Ural Mountains (Stiller *et al*., 2014). Molecular clock analysis has previously been used on estimating divergence times between *Spirometra* and *Diphyllobothriidae* developing robust methodologies for mitochondrial mutation rate occurring within

cestode parasites (Zhang *et al*., 2016). In this study, molecular clock studies will be used alongside molecular p-distance to estimate divergence times between geographically specific isolates of *D. dendriticum.* Using this method can estimate divergence times between European populations, revealing trends in the transplacement of parasite species throughout region. With salmonid populations being frequently trans-located, studying the ancient movement of the parasite may reveal correlations between *D. dendriticum* presence and salmonid transplacement particulary within new world environments such as South America where the parasite is present within populations of *O.mykiss* (Torres *et al.,* 1981; Torres *et al.,* 1993; *Mercado et al*., 2010; Rozas *et al.,* 2012). The study would also reveal the involvement that the eastern movement of carnivorous mammals has had in transplacing the parasite across mainland Europe and the British Isles mirroring trends seen in previously discussed parasites.

In this current study DNA barcoding techniques were used to accurately identify species of *Diphyllobothrium*, and the efficacy of cox1 and ITS2 sequences as markers for the identification of *Diphyllobothrium* was compared to inform future studies regarding species identification of plerocercoid infecting intermediate fish host. The second part of the study will concentrate on the phylogeographical existence of *Diphyllobothrium dendriticum* within the Gairloch region. Haplotype network construction will concentrate on the ascertaining whether *Diphyllobothrium* isolates are geographically specific, with molecular clock analysis seeking to understand the reason for the transplacement and maintenance of endemic infective regions of the species throughout Europe and into South America.

4.2 Materials and methods

4.2.1 Sample collection

A total of 12 plerocercoid cestode larvae isolates were molecularly characterized from dissection of whole brown trout donated by anglers from 2013-2016. Parasites were obtained from 3 different lochs from the Gairloch region of North West Scotland (centered approximately latitude: 57º43’20.59”N, longitude: 5º38’6.85”W). Each

collection site of infected fish were denoted as; Lochan nam Breac (latitude: 57°44’2597”N, longitude: 5°40’2979”W) Loch Feur (latitude: 57°44’20, 13”N, longitude: 5°40’57.7308”W) and Loch Coire nah-Airigh (latitude: 57°44’29.7348”N, longitude: 5°41’26.2464W).

4.2.2 DNA extraction, amplification and assembly of cox1 and ITS2 fragments

Tissue snips were taken from 12 plerocercoid worms. Snips were homogenized in ATL buffer (Qiagen Inc.), total genomic DNA was extracted using protocol of Qiagen™ Blood and Tissue extraction kit. For each specimen, partial *cox*1 fragments were amplified using the cestode-specific PCR primers: BW3 (forward; 5’-TTT TTG GCC ACC CCG AAG TAT AT-3’) and BW4.5 (reverse; 5’-TAG TGA CAT TAC ATA GTG GAA GTG-3’) (Wicht *et al*., 2007) and the ITS2 marker was amplified using the primers of Logan *et al*. (2001) FLO1 (forward; 5’-GAG CGC AGC CAA CTG TGT G-3’) and ITSII (reverse; 5’-CGG TGG ATC ACT CGG CTC-3’).

PCR reactions were performed for each gene fragment using 12.5 µl Thermo–Start^{R}PCR master mix (0.625 Units of Taq DNA polymerase, 1X reaction buffer, 0.2 mM of each dNTP and 1.5 mM $MgCl_2$) and 1–2 ng/µl of DNA. The template PCR program used for all isolates was: 95° for 15 minutes; 40 cycles at 95°C, 1 min; 55-58°C; 72°C, 1 min; 72°C for 7 minutes. Annealing temperatures were customized on the primer set with FLO 1/ITSII using 57°C and BW3/BW4.5 52°C. Final reactions were made up to 25 µl with PCR-grade water. Reactions were performed using a Techne Prime 96 well thermal cycler (Techne™) and 5 µl of each amplicon was visualized in 1 % agarose gels stained with SafeView nucleic acid stain (nbsbiologicals™) under UV using U: Genius Syngene gel documentation system. The remaining 20 µl PCR products were sequenced at the DNA sequencing facility of the Natural History Museum, London, using the PCR primers with Fluorescent Dye Terminator Sequencing Kits (Applied Biosystems™); sequencing reactions were run on an Applied Biosystems™ 3730XL automated sequence

4.2.3 Identification of species and phylogenetic reconstruction

DNA sequences of cox1 and ITS2 were assembled and edited using Geneious v 8.5 (Kearse *et al*., 2012), with BLAST searched being performed through NCBI GenBank sequence database for initial identification of worms (Table 1). The BLAST search also was used to ensure that there was no contamination from other organisms during DNA extraction process. Initial results of BLAST search revealed sequences generated in this study exhibited high similarity to *D. dendriticum.*. To deal with heterozygous sites arising from direct sanger sequencing PHASE analysis package as part of DNAsp software (Librado and Rozas., 2009) was utilized to generate a dataset arising from heterozygote alternative sequences and initial sequence. Prior to phylogenetic reconstruction using MrBayes, ambiguity needs to be set within the original nexus datablock. The discounting of ambiguous sites is done using '{heterozygous base position}' within the data block itself, this requires editing from original full data block generate via Mesquite program (Maddison and Maddison, 2018). Heterozygous sites along with indels were ignored, Indels within nuclear marker were deleted within the alignment, removal is done to ensure that hypervariable or repeated regions do not affect the final phylogenetic signal within nuclear constructed phylogenies.

To construct phylogenetic trees, published sequences were acquired via NCBI Genbank, newly generated *cox1* and ITS2 sequences were aligned using Clustal W implemented in BioEdit software (Hall, 1999).

Phylogenetic tree reconstruction was performed using MrBayes 3.2.6 (Ronquist *et al.*, 2012). Markov Chain Monte Carlo (MCMC) were run for 10 000 000 generations, sampled every 100 generations. Bayesian posterior probability values representing proportion of samples recovering set clades were estimated after the initial 1000 trees were set as burn-in. Final tree was viewed using FigTree 1.4.3 (Rambaut, 2003).

4.2.4 Assessment of molecular diversity

Diversity data was acquired through the use of DNAsp (Librado and Rozas, 2009), for comprehensive analysis of DNA polymorphism data. The genetic marker cox1 has been indicated as a marker that can successfully differentiate between geographical lineages in other species of cestode (Haukisalmi *et al*., 2016; Kinkar *et al*., 2016). The cox1 marker

is utilized provide insights into the evolutionary relationships between populations of *Diphyllobothrium* sampled in this study and previously published Eurasian isolates.

Sequences from cox1 amplifications were aligned using the same methods as described in phylogenetic construction methods. Molecular diversity within the alignment group was calculated using haplotype analytical methods, with number of haplotypes (***h***) and haplotype diversity (**Hd**) being considered within DNAsp (Librado and Rozas, 2009). PoPArt software (Leigh and Bryant, 2015) was used to further visualize haplotype diversity across the *D. dendriticum* dataset through building graphical haplotype networks. To compare molecular characteristics of cox1 and ITS and assess their use as markers for species identification and molecular diversity DnaSP 5 (Librado and Rozas, 2009) software was used. The measures of generic nucleotide diversity were calculated on alignment set including number of segregating site (**S**), nucleotide diversity (**Pi**) and average pair wise divergence (**K**).

4.2.5 Divergence times

Approximate divergence times between isolates of *D. dendriticum* were estimated with an uncorrelated lognormal relaxed-clock model using software BEAST 2.3.6 (Drummond and Rambaut, 2007) (Table 2). Divergence time trees were calibrated using the molecular clock scale with set divergence times based on 0.0225 (Zang *et al*., 2015) substitutions per site per million years. Molecular clock was not rejected as the likelihoods were high and p was not significant (*InL*– 3128.87; $p<1.00$) supported with strong posterior values (ESS; 432). Relaxed molecular clock was used as it showed a greater level of statistical certainty with a higher likelihood (*InL*- 3128.87) and low values of HDP (95% highest probability density).

4.3 Results

4.3.1 Phylogenetic species identification

The cox1 sequences produced an alignment of 389bp without gaps. ITS2 sequences produced an alignment of 470bp editing was performed to compensate for length variation and potential sequencing error. Phylogenetic reconstruction produced similar topologies within cox1 (fig 12A) and ITS2 alignments (fig 12B). Newly derived cox1

isolates from Gairloch were all placed alongside *D. dendriticum* samples from Russia (JQ245479-JQ245481) and Chile (JN152993-JN153005) (fig 12A). All published samples within the *D. dendriticum* clade derived from plerocercoid stage infecting salmonid intermediate hosts. In total the cox1 indicated three clades within *D. dendriticum* populations. As discussed in Kuchta *et al*.'s previous (2013) phylogenetic construction of *Diphyllobothrium* species trees cox1 is unable to delineate Chilean *D. dendriticum* samples (JN153004, JN153996, JN153995, JN153994, JN153993, JN153005) and Canadian *D. ursi* (AB605762, AB605763), forming a distinct lineage apart from the remaining *D. dendriticum* clade.

Regarding the phylogenetic relationship of *D. dendriticum* using cox1 marker (fig 12A), samples fall into four sub clades. Sub clade 1 consists of sequences deriving from Holland, Czech Republic, Switzerland and Lake Baikal in Russia. Sub clade 2 represented sequences deriving from plerocercoid infective stages from Chilean salmonids, with the third large clade deriving from the Gairloch study area. The fourth clade represented mixed geographical sample locations, with isolates deriving from Scandinavian, Russian and Scottish fish host populations. There was a clear division between Eurasian and South American isolates of *D. dendriticum,* however no clear country-specific geographical relationship within Eurasian clade.

Phylogenetic construction deriving from ITS2 (fig 12B) sequences corroborated with the placement of Gairloch samples alongside samples of *D. dendriticum*. Unlike the *cox*1 species tree, ITS2 indicated an overall lack of species delineation of *D. dendriticum* and *D. ditreum* (AY549515, AY549514 and DQ386128) isolates. The phylogenetic reconstruction also demonstrated substantially lower bootstrap values and branch lengths.

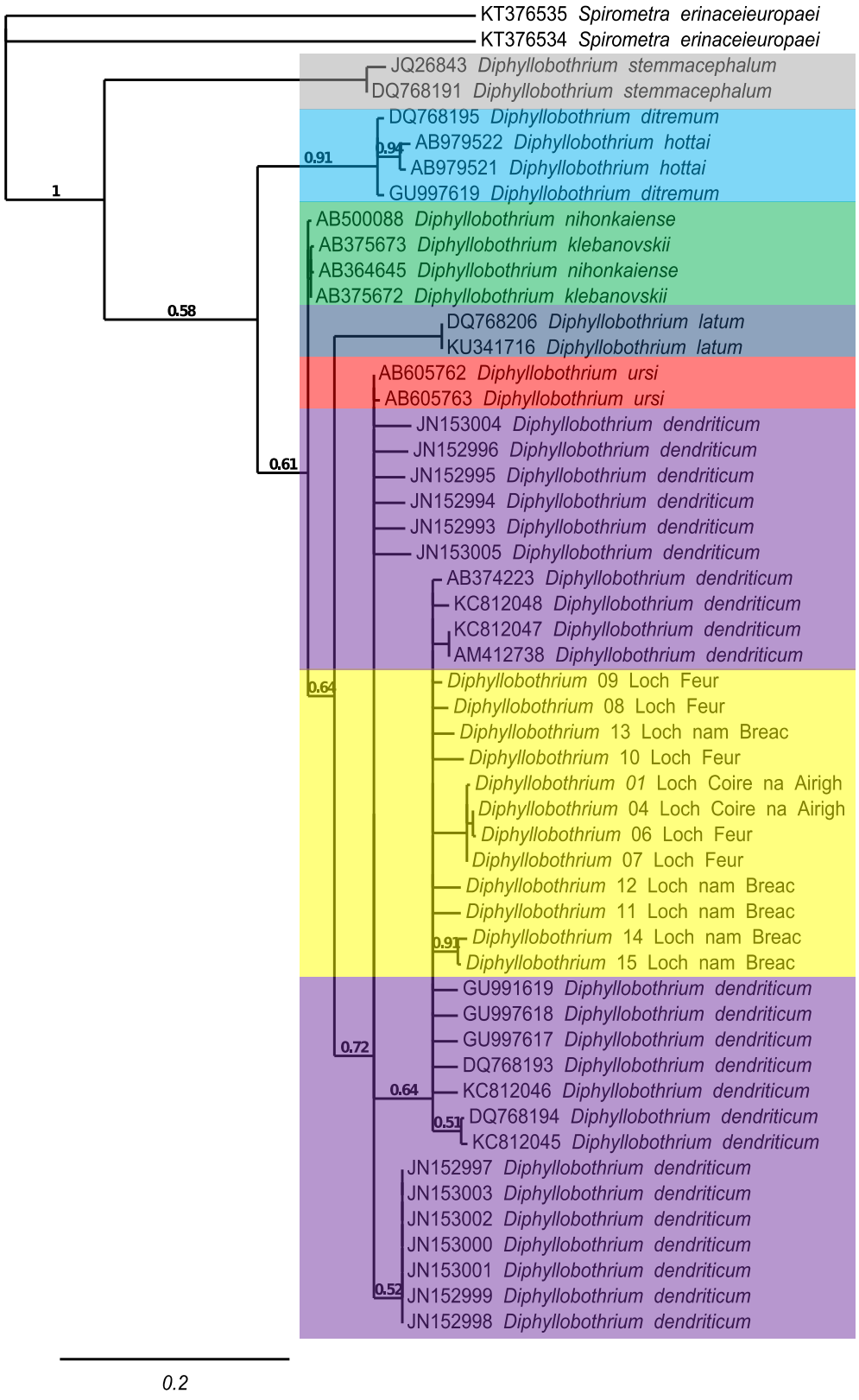


Figure 13; Phylogenetic reconstruction of genus *Diphyllobothrium* utilizing mitochondrial *cox*1 DNA barcoding marker and *D. dendriticum* only tree using *cox1* DNA barcoding marker. Worldwide tree (A) was constructed using TN93 model conditions; tree indicated a placement of Gairloch study-derived samples alongside *D. dendriticum* isolates. *D.semmacephalum* clade is highlighted in grey, D.ditreum/*D.latum* clade is highlighted in blue, *D.nihonkaiense* and *D.klebanovskii* clade is highlighted in green *D.latum* clade is highlighted in navy blue *D.dendriticum* is highlighted in purple Gairloch derived isolates are highlighted in yellow.

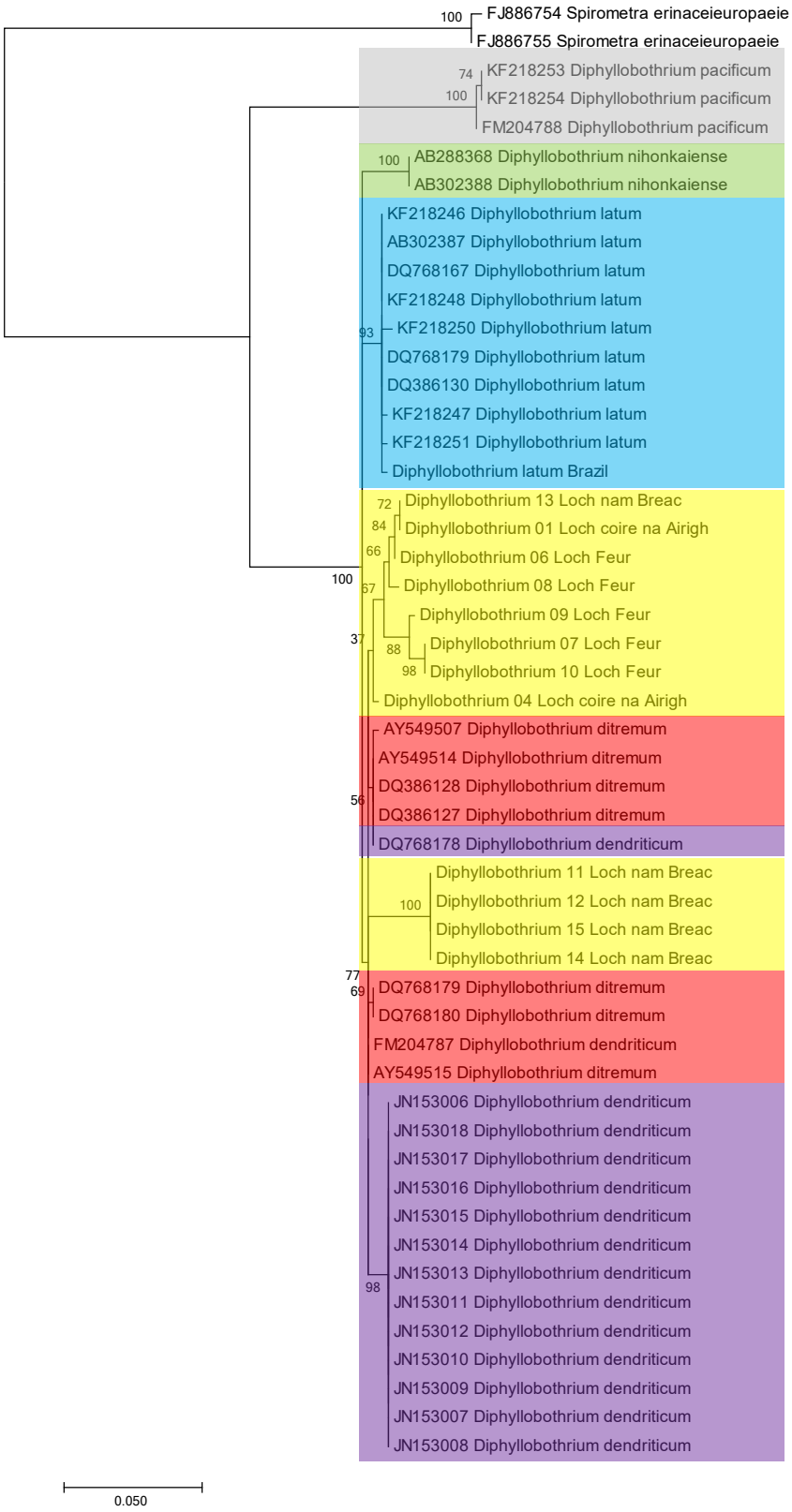


Figure 14; Phylogenetic construction of *D. dendriticum*-only isolates were created using HKY model conditions and showed a distinct Gairloch isolate-specific clade of *D. dendriticum*. Phylogenetic construction of *Diphyllobothrium* genus using nuclear ITS marker under the TN93 model conditions. *D. pacificum* is highlighted in grey, *D. latum* is highlighted in blue, *D.ditreum* is highlighted in red, *D.dendriticum* is highlighted in purple and Gairloch derived isolates are highlighted in yellow.

4.3.2 Evolutionary relationships between geographical isolates of *D. dendriticum*

Cox1 alignments consistently displayed greater nucleotide diversity and divergence across 13 plerocercoid samples within the Gairloch loch system compared to samples from mainland Russia, Norway and Chile (Table 7). The cox1 Gairloch isolates displayed increased genetic diversity relative to all other clades with the 16 samples being represented via 11 haplotypes and a haplotype diversity value of 0.985 (Table 7A). Pooled samples from European isolates in Holland, Czech Republic and Switzerland represent the mainland Europe clade, demonstrated distinctively higher genetic diversity-related values with pi; 0.01688 and 10 segregating sites from only 3 samples. Gairloch isolates demonstrating considerably higher molecular diversity (S; 10 and pi; 0.13) and divergence (K; 5.13636). Geographical isolates showing least genetic diversity derived samples from Lake Baikal, these showed no genetic diversity sharing one haplotype between the 3 samples (Figure 13).

A

Site	S	Pi	K	h	Hd
Chile	6	0.00331	1.30769	5	0.692
Scotland	2	0.00517	2	2	1
Norway	2	0.00506	2	2	1
Russia	0	0	0	1	0
Gairloch	16	0.013	5.13636	11	0.985
Europe	10	0.01688	6.666	2	0.667

B

	1	2	3	4	5
Chile					
Europe	0.052504				
Gairloch	0.050505	0.03588			
Scotland	0.052083	0.036616	0.026199		
Norway	0.044508	0.02904	0.020833	0.007576	
Russia	0.043624	0.028283	0.021023	0.012879	0.008333

Table 6; A Diversity indices across worldwide isolates of *D. dendritcum* S; segregating sites between *cox1* genes, Pi pairwise diversity, K average diversity within sub set, h; number of haplotypes in sub set and Hd; pairwise diversity between sub set haplotypes. **B** P-distance between geographic isolate sub-sets in Chile, Loch Lomond, Gairloch, Scotland (other), Norway and Russia.

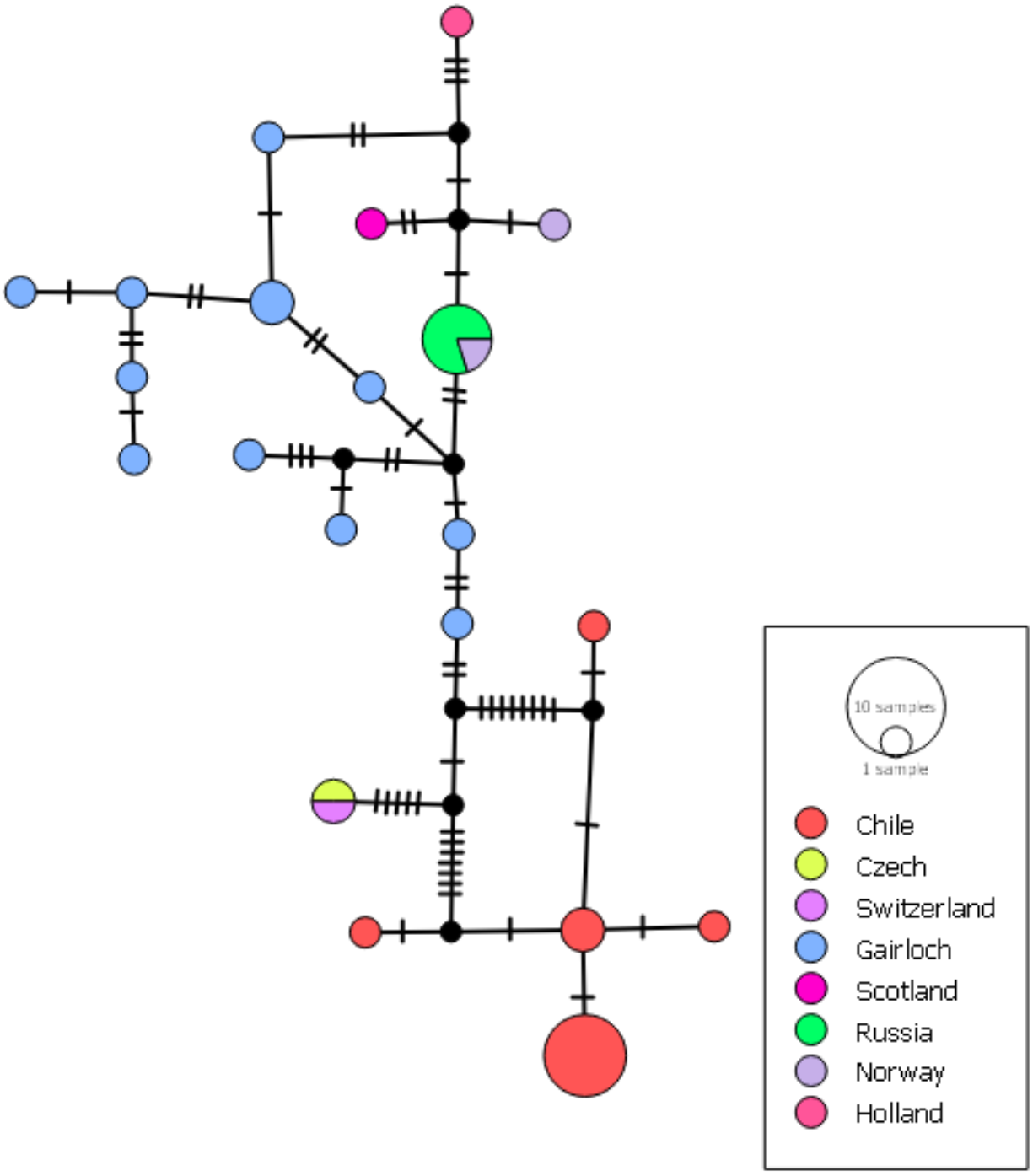


Figure 15; Haplotype network analysis constructed using *cox1* DNA barcode data indicating geographic-specific haplotypes within Chilean, Gairloch, Scotland and Dutch geographic isolates. Shared isolates were only found between isolates found in the Czech Republic and Switzerland and between Russian and Norway- derived isolates.

4.3.3 Estimation of divergence time between Scottish and European lineages

As shown in Fig 13 geographical isolates of *D. dendriticum* reveal numerous temporal segregations in response to relative geographical isolation. The early branching of Chilean isolates (JN152993-JN152005) from Eurasian samples have an estimated most recent common ancestor (TMRCA) of 0.9874 (SD 0.00833) indicating a divergence from Eurasian isolates 987,400 years ago (Table 8, Figure 14). The large Eurasian clade reveals distinct division from mainland European isolates and isolates deriving from Scotland, with TMRCA of the two clades being 0.3821 indicating an approximate divergence time of Scottish isolates from mainland European occurring 382,100 (SD.00479) years ago. Within the mainland European clade divergence between clade including Norwegian and Russian isolates and central European clade including samples from Czech republic, Switzerland and Holland revealed to have occurred 498,100 years ago (TMRCA 0.4981; SD 0.00331). The molecular clock analysis implemented in BEAST indicated that a relaxed molecular clock provided a more reliable estimation of divergence between clades when compared with the output of strict molecular clocks under Yule speciation model conditions, molecular clock calibrations were not rejected due to high likelihood values and non-significant *p* values (InL:–3128.87; *p :* 1.00). The ESS values (table 2) for each TMRCA was consistently higher under the relaxed clock conditions in comparison to strict molecular clocks, therefore data deriving from strict molecular clock was rejected in favor of relaxed model outputs.

Parameter	*N*	Mean (SD)	ESS	95% HPD (upper/lower)
Likelihood (total)	35	-907.3573	845.998	-1814.525
TMRCA (Chilean)	12	0.9874 (SD0.00833)	708.332	1.2231/0.7517
TMRCA (Norway/Russia)	7	0.4981 (SD0.00331)	900.886	0.5832/0.413
TMRCA (Holland/Swiss/Czech republic)	3	0.3481 (SD0.00212)	887.549	0.4474/0.2488
TMRCA (Scotland/Gairloch)	13	0.3821 (SD0.00479)	901	0.462/0.3022

Table 7; Parameters separating lineage ancestor groupings of different *D. dendriticum* isolates. Mean ± SD; likelihood divergence parameters, scaled up (10^3) to achieve TMRCA (The most recent common ancestor) between clades of molecular clock tree. ESS; Effective sample size, effective sample size of trees sampled in Markov chain posterior analysis 95% HPD; indicates 95% confidence interval that exists according to variance.

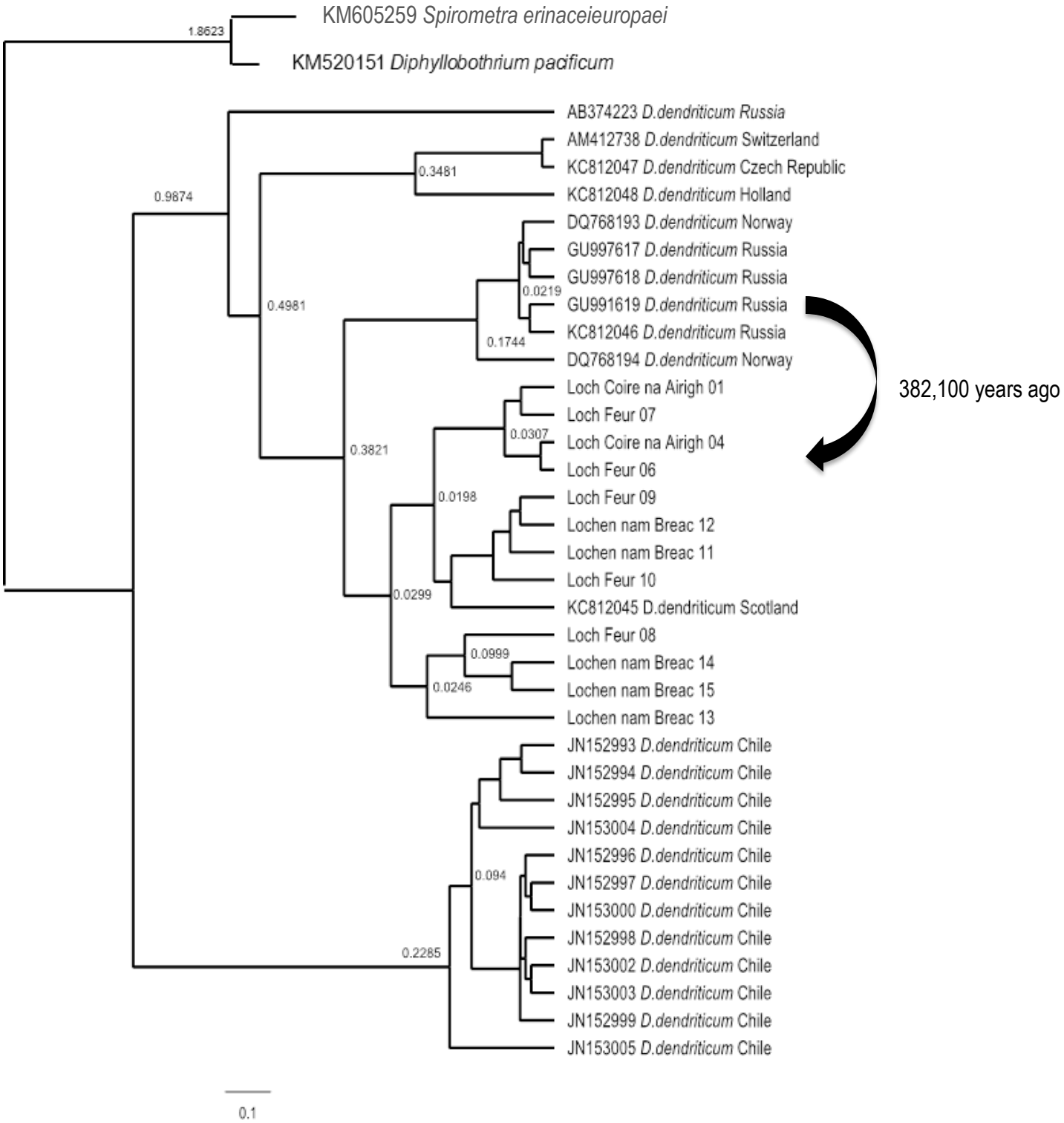


Figure 16; Maximum Likelihood tree calibrated with the use of estimated cestode evolutionary rates of 0.0225 substitutions per MYA (Zhang *et al*., 2015) in *cox1* in order to assess time of divergence between isolates deriving from geographic localities. The *cox1* phylogeny was constructed using MrBayes, with molecular clock performed on phylogeny using BEAST v2 and was shown to be evolving in a clock-like progression under the hypothesis testing values of (*InL* – 3128.87; *p* : 1.00). There appear to be three main clades within *D.dendriticum* one which contains exclusively Chilean isolates, one including exclusively Scottish isolates and finally one mainland European clade including isolates from Holland, Switzerland, Czech republic, Norway and Russia. Initial divergence appears to have occurred between a basal Eurasian clade and South American clade with further divergence of central European and Scottish isolates from Russian/Scandinavian isolates.

4.4 Discussion

An important outcome in the investigation was the confirmation that *Diphyllobothrium dendriticum* was present within the populations of *S.trutta* studied. This conclusion was reached in conjunction with the placement of Gairloch-derived samples alongside *D. dendriticum* samples in phylogenetic construction of both cox1 and ITS2 species trees. The ITS region however did present as problematic marker to be used for accurate identification of *D. dendriticum*. Having been highlighted as a conserved genomic target, ITS2 markers seemed to lack the ability to discern clearly between *D. ditreum* and *D. dendriticum* (Skerikova *et al*., 2006). Analysis of *cox1* genetic markers proved to be most accurate at species level identification, providing differentiation between *D. ditreum* and *D. dendriticum* isolates. Nuclear markers proved inconclusive due to the inability to differentiate between *D. ditreum* sequences (DQ768180, DQ768179, AY549507, AY549515 and DQ386128), sequences derived from current study and Chilean isolates (JN153006-KN153015). The ability to differentiate between both species is of high importance because *D. dendriticum* is human pathogenic; both species share the same geographic areas of infection (Holarctic) and intermediate fish host (coregonid and salmonid), increasing the likelihood of intermediate host infection in the same water body. The construction of mitochondrial *cox1* tree indicated the placement of *D. ursi* alongside Chilean isolates, bringing into question the correct identification of *D. ursi* being *D. dendriticum* or Chilean samples of *D. dendriticum* being misidentified as *D. ursi*. This issue has been highlighted in previous phylogenetic tree construction (Scholz *et al*., 2009). The high polymorphism of mitochondrial DNA is well known among species of cestodes (Miyadera *et al*., 2001; Van Steenkiste *et al*., 2015) with nuclear DNA markers exhibiting a more conserved target for identification.

Several studies have stressed the importance that molecular taxonomy plays in delineation of species of *Diphyllobothrium* genus (Yera *et al*., 2006; Wicht *et al*., 2007; Rozas *et al*., 2012), however there is a lack of consensus regarding the correct marker to employ. Investigations have included the use of ITS2 (Logan *et al*., 2004; Skerikova *et al*., 2006), with the difficulty of using ITS2 sequences for phylogenetic analysis primarily attributed to the low levels of variation between the closely- related species of *D. dendriticum* and *D. ditreum*. The *cox1* markers show greater discrimination between

closely related species of *Diphyllobothrium* more effectively than using ITS2 (Year *et al*., 2006). This characteristic has been consistently noted within cestode species (Villas *et al*., 2005) with studies indicating a preference for *cox1* identification of plerocercoid and proglottid stages of the *Diphyllobothrium* life cycle. Thus, because both *cox1* and ITS2 markers were easily amplified from parasite tissue, based on the nodal support, phylogenetic tree reconstruction and uncorrected *p*-distance estimates of divergence, *cox1* should be preferably used over ITS2.

Phylogenetic tree construction revealed three divergent clades between South American, Northern Europe/Russian and mainland Europe/Gairloch isolates. Intra-clade distance analysis revealed that sites within mainland Europe differed by 3.5% and that the extended geographic distance between Chilean isolates and isolates from mainland Europe does not correlate with large genetic differentiation with only a 5.2% distance observed between mainland European clade and Chilean clade of *D. dendriticum*. Without knowing specific locations from which Chilean stocked trout were derived, it seems likely that movement of salmonids to South America has trans-placed European haplotypes of *D. dendriticum.* From phylogenetic analysis it is clear that Chilean *Diphyllobothrium* isolates are genetically divergent from its European ancestors, however, not to the extent that they constitute a different South American species.

Observations of the Scottish clade of *Diplyllobothrium* dendriticum through phylogenetic and haplotype analysis make it clear that Gairloch isolates are genetically divergent from mainland European isolates. With Scotland no longer supporting populations the required definitive host carnivorous mammals, the historical movement of definitive host species may explain the parasites presence in the Gairloch system. Definitive host movement can explain the presence of many species of parasites in select habitats and localities (Kennedy, 1993; Gortazar *et al*., 2007). In the case of *Diphyllobothrium* the movement of the definitive Ursid (Bear) definitive host (Frechette and Rau, 1978; Yamasaki *et al*., 2012) is the most likely host involved in the trans-placement of the parasite through-out Europe (Loreille *et al*., 2001; Taberlet and Bouvet, 1994). Divergence dating indicates that an ancient lineage is present of *Diphyllobothrium* within mainland Europe with

divergence between European and Scottish isolates being potentially associated with latitudinal movement of ursid species (Valdiosera *et al*., 2007; Markova *et al*., 2010). Ice sheet formation and recession during Pleistocene glacial period would allow for isolation of Ursid species within the British Isles with evidence indicating potential chalk bridge collapse residing in the Dover straight 400,000 years ago (Gupta *et al* 2017). Along with presence within Europe there is also a clear ancient lineage of *D. dendriticum* within Chilean freshwater fish, with salmonids being trans-located to Chile in the 1900's the ancient presence of the tapeworm within the region is most likely related to infected definitive host presence within South America with the native *Tremarctos ornatus* (Spectacled bear) being a potential carrier maintaining the infection within Chilean native freshwater fish species (Torres *et al.,* 2004) before the arrival of salmonids in the late 1900's (Krause *et al*., 2008). Whilst molecular clocks provide potential reasoning associated with parasite infection spread across regions, the timeframe and potential variables make any conclusions somewhat tentative. Also due to the analysis done purely using mutation times of cestodes (Zhang *et al*., 2016) rather than the use of fossil data, the analysis provides a snapshot of the potential trans placement of *Diphyllobthorium dendriticum* within the UK rather than provide an in-depth systemeatic review of the evolution of the species. Increased field sampling methods with additional definitive and freshwater host sampling could provide evidence to establish more concrete conclusions.

4.5 Conclusion

Molecular identification of parasite species is the only truly reliable method of species identification as cestode morphology continues to prove to be problematic in species delineation. Molecular approaches present an extremely powerful tool in fish health analysis and the identification of human infective species, however a detailed understanding of the taxonomic and phylogenetic construction power different molecular markers present and which marker presents the best option to use. As with other freshwater cestode species mitochondrial *cox1* markers and nuclear ITS2 markers have the ability to discern between species of *Diphyllobothrium, with cox1* proving to be more accurate in phylogenetic species delineation. The use of *cox1* marker is shown to be useful for providing tentative evolutionary insights into the population biology and

historical movements of *Diphyllobothrium.* Molecular approaches to accurate taxonomic identification of medically and veterinary important parasite species can be extremely important in monitoring infection dynamics and possible control of food-borne tapeworm-associated disease. Currently there are few data regarding *Diphyllobothrium* presence in the UK, and a limited understanding of the species in Eurasia (the majority of isolates collected within Europe are from within central European alpine regions and Scandinavia). Consequently, only through further molecular-based studies of *Diphyllobothrium* will a more complete story of the broad fish tapeworm be achieved.

Thus, within this chapter and the previous chapter use of multiple genetic markers has highlighted the importance of proper methodology in using molecular techniques to identify parasite species in salmonids. Moreover, although it has emphasized the importance of genetic identification because of the pathogenic nature of some parasites within humans, it also permits elucidation of potential historical movement of the species throughout Europe.

Brown trout within the Gairloch system exist under parasite infective insult with a varied milieu of parasite infection throughout isolated host populations. With brown trout surviving in a diverse infection environment, the host's immune system is under a constant and varied attack throughout the system, with a clear dichotomy between single parasite infection populations and isolated populations with a multitude of parasite infections. This division in differential parasitism makes the host populations a model system for the investigation of evolutionary and functional effects differential parasitism presents to a host immune system within a freshwater environment.

5. Role of parasite-driven selection maintaining MHC diversity within wild populations of UK salmonids

5.1 Introduction

The adaptation of salmonid fish across geographic ranges occurs in response to local selective pressures acting to maximize population fitness within local environment (D J Fraser et al., 2011; Junge et al., 2011; Primmer, 2011; Gharrett et al., 2013). Fish populations become locally adapted due to a variety of selective pressures; differential hydrology (Crossin *et al.*, 2004), change in food resource (Jonassen *et al.*, 2000), predator presence (Cousyn *et al.*, 2001) and parasite insult (Bernatchez and Landry, 2003). Parasites are considered to be among the strongest selective force acting on population (Lamaze *et al*., 2014b; Seifertová *et al*., 2016; Zueva *et al*., 2014). The ease of which host populations can adapt to a novel parasite infection is crucial, to survive hosts must acquire resistance (Lymbery and Thompson, 2012; C Monzón-Argüello *et al*., 2014), or undergo possible extinction (Thieltges *et al*., 2009).

One key host response is immune gene adaptation to parasites within the immediate environment (Sommer, 2005; Zueva *et al*., 2014) and a vital set of genes involved in this adaptation of response is the major histocompatibility complex (MHC). MHC molecules are expressed in two different classes; MHC class I acts as an intracellular immune response initiator recognizing viral or bacterial infections whilst MHC class II recognizes extracellular proteins (Simpson, 1988). The recognition site of the MHC II includes a "basket" receptor denoted the "antigen binding site" (ABS) (Cuesta *et al*., 2006; Cutrera *et al*., 2014; Forsberg *et al*., 2007; Medel *et al*., 2010)which directly interacts with parasite proteins (Figure 15) (Eizaguirre *et al*., 2012; Lamaze *et al*., 2014a; Natsopoulou, 2010). Genes that encode MHC II are some of the most polymorphic genes in vertebrates with increased variability within the antigen binding site (Schenekar and Weiss, 2017). Genetic theory suggests that at population level the resistance/tolerance of an array of parasites is associated with high MHC II diversity (Sommer, 2005).A large

body of empirical data that varied pathogen presence acts on MHC loci maintaining MHC diversity within populations (Radwan *et al*., 2014; Bracamonte *et al*., 2015: Perchouskova *et al*., 2015; Schuster *et al*., 2016). Recent studies suggested positive association between MHC II diversity and parasite diversity within human (Prugnolle *et al*., 2005) and salmonid (Dionne *et al.,* 2007) models.

Due to salmonids' varying conservation status worldwide (Waples and Hendry, 2008) and increasing reliance on aquaculture, salmonids have been targeted in numerous investigations regarding health and welfare of wild and farmed populations (Grimholt *et al*., 2003; Zhang *et al*., 2015b). To advance the understanding of fish health, numerous investigations have taken aim at molecular processes underlying infection resistance and susceptibility (Grimholt *et al*., 2003; Mjaaland *et al*., 2005; Dionne *et al*., 2009; Consuegra and Garcia de Leaniz, 2008; Zueva *et al*., 2014). Genetic diversity of salmonid MHC II complex attributes to increased population fitness (Dionne *et al*., 2007; Lamaze *et al*., 2014a; Miller and Withler, 2004). The diversity of MHC II is linked to a number of different factors such as; mate choice (Consuegra and Garcia de Leaniz, 2008), embryo and juvenile fitness viability (Forsberg *et al*., 2007) and population history of pathogen infections (Lamaze *et al*., 2014a).

Genetic variation of salmonid MHC II genes is an evolutionary response to parasite infection and contributor to increased fitness of fish populations (Buchmann and Uldal, 1997; D. J. Fraser et al., 2011; Lamaze et al., 2014a). Two main hypotheses have been used to explain parasite-driven selection; the Over-dominance hypothesis (Penn, 2002) and Negative frequency dependent selection hypothesis (Piertney and Oliver, 2006). The over-dominance hypothesis proposes that heterozygous individuals exhibit greater fitness due to increased tolerance to a broader array of parasite antigens compared to homozygous individuals (Hansson and Westerberg, 2002). The negative frequency-dependent selection hypothesis states that rare host MHC-associated alleles are advantageous due to a limited co-evolutionary response of parasite (Koskella and Lively, 2009). Empirical data supporting the specific role that either process play in wild populations is currently lacking (Apanius, 1997). A wide body of research based on parasite-driven selection in salmonids has been aimed at MHC associations in; anaemia

virus (Mjaaland *et al*., 2005), *Aeromonas salmoncida* bacteria (Croisetière *et al*., 2008), *Gyrodactylus salaris* (Malberg, 1957; Tonteri *et al*., 2010) and stocked vs wild population adaptation (O'Farrell et al., 2013a; Schenekar and Weiss, 2017). There is a paucity of studies based on parasite-driven selection within wild salmonid populations, with methodologies currently based on hypothesized parasite presence (Cohen, 2002; D. J. Fraser et al., 2011) or comparisons within immunogenically naïve stocked fish (Byrne *et al*., 2002; Milot *et al*., 2013; Schenekar and Weiss, 2017).

Figure 17; 3D structure of MHC II protein, purple and green structure denote Antigen binding site with yellow structure denoting pathogen protein (Castellino, 1997)

Populations of brown trout have survived in the northern Highlands of Scotland since the retreat of last Scottish ice sheet (Late Devensian; 17,000 – 13,000 years ago) (McKeown *et al*., 2010). Numerous fish health screens of Scottish brown trout indicated high levels of parasite infection, with the species acting as an intermediate host in various cestode, nematode and trematode life cycles (Byrne *et al*., 2002; Hartvigsen and Kennedy, 1993; Molloy *et al*., 2014). Brown trout in the Gairloch region of Scotland inhabit an assortment of isolated freshwater hill lochs and sub-alpine streams. Unlike many populations of Eurasian brown trout, freshwater bodies in this region have not undergone population expansion through non-native fish stocking or fish movement (Dick *et al*., 1987; Ferguson, 1989; Byrne *et al*., 2002; Kohout, 2013). With host sub-populations undergoing long-term exposure to local parasite infection the region is ideal for investigating effects that parasite-driven selection has upon host immunogenetics. Previous parasite screening (Chapter 2) indicates that Gairloch trout populations show parasite taxon diversity inversely associated with trout habitat isolation.

5.1.1 Aims and Objectives

This study sequenced specific MHC class II locus (*Satr*-DAB in *S.trutta*) in 14 wild Scottish brown trout sub-populations to assess the adaptive genetic variation in response to differential parasite infection throughout each sub population. Choices of populations were based upon parasite infection diversity of host individuals. Key subsets were; Elevated lochs, only single parasite species infecting brown trout and Western lochs, where hosts were infected by a variety of parasite species. The study addressed three specific questions:

1. Does immunogenetic MHC II variation exist within the Gairloch system within brown trout populations?
2. Does spatial genetic population structuring occur within the *Satr*-DAB gene in the Gairloch system brown trout populations?
3. Does parasite infection exert selective pressure on the MHC II to maintain genetic diversity in the *Satr*-DAB gene within these populations?

Study data will provide insights into the role that parasite infection plays as a selective pressure in wild populations and, additionally, may increase ecological benefit and sustainability of future captive breeding strategies and conservation measures associated with socioeconomically important salmonid species.

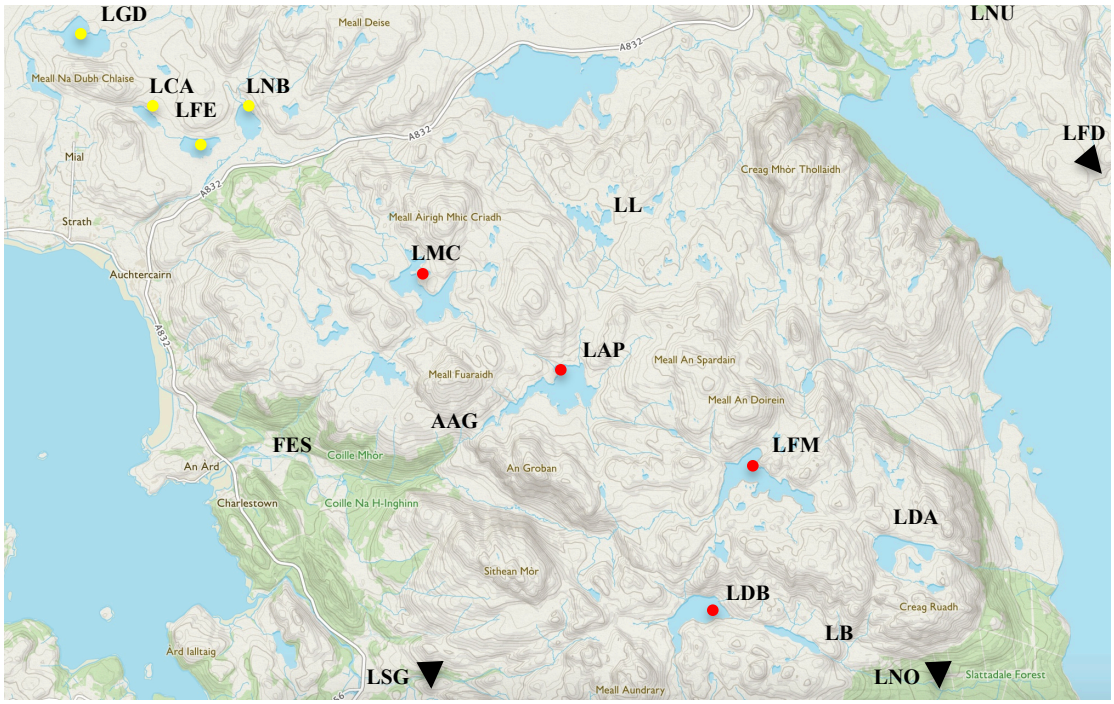


Figure 18; Map of sampling regions within Gairloch system abbreviations are used for specific loch sub-populations: LNO; Loch na h-Oidche, LB; Loch Buinachien, FES; Flowerdale estuary, AAG; Alt a'Glinne, LFM; Loch na Feithe Mugaig, LDA; Loch Doire nah-Arigh, LAP; Loch Arigh a'Phuil, LMC; Loch Mhic Criadh, LL; Loch Laraig, LFE; Loch Feur, LNB; Lochan nam Breac, LCA; Loch Coire na h-Arigh, LGD: Loch Gharbe Doire, LNU; Nursery Lochan and LFD; Loch Fada. Division of elevated and western loch populations are done with red dots for elevated loch populations and yellow for western loch populations. (Google maps, 2018)

5.2 Methodology

5.2.1 Sample populations and parasite screen

Whole fish samples were collected from 2011-2014, via angler donation in conjunction with the Wester-Ross Fisheries Trust. Tissue extracted from 210 whole Brown trout was used for the PCR amplification of *Satr*-DAB gene. Tissue samples were taken from internal musculature to ensure no cross contamination whilst under direct contact during field sample study conditions, and were stored within individual 1.5ml microcentrifuge tubes within 70% ethanol before DNA extraction (Figure 16).

Parasite acquisition was via full internal parasite screen across trout obtained from all sampled populations and was according to fish health screen protocol (Environment agency, 2011). One key aspect of sample population segregation is done using parasite screen data deriving from previous fish health screen analysis (Chapter 2). To investigate the potential effects of parasite-driven selection those trout screened for parasitic infection were divided into two sub-sets; those infected by only one parasite taxon and those infected by more than one. In the event, this division was also reflected in their geographic distribution; those with only one parasite taxon tended to be located in 'elevated' lochs (eastern, high altitude) and those with more than one parasite taxon located in the 'western' lochs (western, low altitude)

5.2.2 MHC II βsequencing

The initial step was to amplify the specific antigen-binding site (ABS) of the MHC II encoded by exon 2 of the MHC II gene which exists as a single copy in salmonids. Total genomic DNA was extracted using protocol of Qiagen^{TM} Blood and Tissue extraction kit. A 252 bp fragment of exon 2 of the MHC II locus (Jacob *et al*., 2010), was amplified and sequenced. The primers used were CL007 5'-GAT CTG TAT TAT GTT TTC CTT CCA G-3' (Olsen *et al*., 1998) and the self-designed reverse primer 5'-CAC CTG TCT TGT CCA GTA TG-3'. PCR reactions were performed for each gene fragment using 12.5 µl hot start Taq polymerase; Thermo–Start^{R}PCR master mix (0.625 Units of Taq DNA polymerase, 1X reaction buffer, 0.2 mM of each dNTP and 1.5 mM $MgCl_2$), 5 ng/µl of DNA and 3.5 ng/µl of each primer. The template PCR program used for all isolates was:

95° for 15 minutes; 40 cycles at 95°C, 1 min; 55°C; 72°C, 1 min; 72°C for 7 minutes. Annealing temperatures were optimized on the primer set as 52°C.

5.2.3 Selection and recombination of the *MHC II β*

The most commonly applied methods for detecting the long-term effect of positive selection is the *dn/ds* ratio (Hughes and Nei 1988). The rationale of this ratio test is that synonymous mutations are essentially neutral because they do not result in an amino acid replacement within the translated protein. Nonsynonomous mutations that result in amino acid change are more likely to affect selection via expressed polypeptide changes. Thus if the there are situations where mutations are deleterious the *dn/ds* would be >1. If, however, positive selection is significant then the *dn/ds* ratio is <1.Selection occurring on the protein coding level of the *Satr*-DAB was evaluated using the ratio of nonsynonymous substitutions per site (dN) to the substitutions per synonymous site (dS). This ratio was calculated using DnaSP software (Rozas, 2009), using Jukes-Cantor correction ratio for potential multiple hits. A codon-based Z-test was implicated using Hyphy population genetics software (Pond and Frost, 2005) to test for positive selection on codons, using 1000 bootstrap replicates for accuracy. The estimation of individual codon-specific dN/dS ratio was performed using program CODEML that is part of the PAML 4.7 bioinformatics software package (Yang 2007). The MHC within salmonids is polymorphic, with additional allele analysis needed to reconstruct haplotype data from raw chromatograph data; this was performed using PHASE haplotype analysis software (Stephens and Scheet, 2003). Further allelic analysis of expected and observed heterozygosity was performed using ARLEQUIN software (Excoffier *et al*., 2005).

5.2.4 Phylogenetic analysis

Phylogenetic tree construction was utilized to investigate potential presence of geographic lineage sorting throughout the Gairloch system. Construction of phylogenetic trees was performed using Mrbayes Bayesian analysis software with additional model selection using software add-on bModeltest within the software suite. The likelihood

settings corresponded to the HKY + I model with the parameter values estimated from the data; priors were set to default values. Four Markov chain Monte Carlo were run for 1 * 106 generations and sampled every 1000 generations. The first 10% of trees were discarded as burn-in, resulting in 900 sampled trees. To calculate the posterior probability of each bipartition, the majority-rule consensus tree was computed from these 900 sampled trees.

As part of testing the efficacy of phylogenetic analysis, a test for saturation will be carried out on the alignment of *Satr*-DAB. With the gene being highly polymorphic in nature, they can succumb to mutational saturation which would affect the reliability of the phylogenetic reconstruction. To perform test for the presence of mutational saturation within the gene dataset DAMBE analysis software is used (Xia and Xie, 2001) to map phylogenetic distance vs number of transitions and tranversions per polymorphic site within the alignment.

5.2.5 Demographic variation and correlation of spatial variation with *Satr*-DAB population structuring

Historical movement of populations can lead to expansion events within sub-populations leading to an uneven distribution of genetic diversity. This uneven mismatch distribution can skew potential genetic diversity across populations. In the case of this study it is used to provide evidence of presence or absence of parasite-driven selection occurring within the Gairloch system. To infer potential historical population expansion events, the measure of pairwise genetic difference distribution analysis is conducted across *Satr*-DAB isolate dataset. This mis-match distribution analysis was performed using DnaSP v5 software, which plots the frequency of pairwise differences between observed and expected mismatch distribution. This calculation is performed over 1000 iterations, with the raggedness (*r*) statistic being an eventual measure to support any significant departure from expected unimodality. The mismatch distribution analysis is supported by further Fu's *F* and Tajima's *D*, also calculated using DnaSP v5, both calculations are used to

support outcome with *P* value used to identify any significant departure from expected neutrality.

A Mantel test was performed to investigate potential correlation between geographic separation and *Satr*-DAB diversity. A Mantel test is analysis to investigate spatial genetic population structuring. This is used to compare genetic distance (p-distance), Fst, and nonsynonomous mutations in relation with geographic distancing between the populations. The Mantel test was performed using APE: Analyses evolution package (Paradis et al., 2004) as part of the R statistical software suite; R-3.5.2 (Ihaka and Gentleman, 1996). To generate slope to calculate potential significance of relationship linear model within R package was used with *n* of repetitions set to default (999), test for significance was based on a null hypothesis of 0.01 significance (Smouse et al., 1986).

5.2.6 Host-parasite diversity correlation

Because *Diplostomum* is the most prevalent parasite within the Gairloch system it is likely to exert a significant selective pressure on MHC II of *S.trutta* hosts. Not only is *Diplostomum* the most prevalent parasite, it is the only one infecting most sub-populations within the overall dataset. In order to measure potential selective pressure exerted by genetic diversity of *Diplostomum* acting on *Satr*-DAB, nucleotide diversity (pi) of *cox1* and ITS markers from 7 different sub-populations of *Diplostomum* were analyzed alongside genetic diversity (pi) of *Satr*-DAB. To infer any possible relationship existing between parasite genetic diversity and *Satr*-DAB diversity, regression analysis was performed using genetic diversity data (pi) deriving from previous *Diplostomum* molecular analysis (Chapter 3) and *Satr*-DAB analysis across select lochs populations where sufficient *Diplostomum* molecular data was obtained. Genetic data was collected from *cox1* markers for *D. baeri* and regression analysis was performed using Minitab v13 statistical software package. Four sample populations of Diplostomum were used were LAM, LNO, LFM and LDA, these sample populations were used due to the Diplostomum being the only parasite found within the populations during parasite screen (Chapter 2).

5.2.7 Satr-DAB visualization

To fully assess the importance of genetic diversity within *Satr*-DAB the implications of the positive selection and genetic diversity on the final amino structure is important. With the antigen-binding site being the primary interaction site between MHC II and parasite protein any conformational or structural changes could affect binding potential between the two proteins. Thus, 3D modelling was used to investigate the potential impact of amino sequence variation on the final ABS protein structure and highlight differential structural factors in response to differential parasite-driven selection within western and elevated loch datasets. The reference *Satr-DAB* LFD 13 was submitted to SWISS-MODEL (Biasini *et al* ., 2003). The model used was murine MHC II with which *Satr-DAB LFD 13* showed 59% similarity. The returned Protein Data Bank (PDB) files were loaded into the supplied SPDV DEEPVIEW (Guex and Peitsch, 1997) program for three-dimensional visualization, graphical manipulations, and the plotting of codons under different selective pressures. SPDV DEEPVIEW was used to output files for the rendering software POV-RAY (Plachetka, 1998), which produces very high quality graphics of the protein. Consensus amino profile divided into two separate groups for means of selection profiling. The first denoted consensus sequences derived from 'singular' parasite infection populations LAM; Loch Arigh mi-criadh, LAP; Loch arigh a'Phuill, LDA; Loch doire na-h arigh and LFM; Loch feithe mugaig. The other group used for comparison of dN/dS data was denoted 'Western', these consisted of LNB; Lochan nam Breac, LCA; Loch Coire na-h Arigh, LGD; Loch Gharbe Doire and LFE; Loch Feur.

5.3 Results

5.3.1 Diversity of *S.trutta* MHC II in the Highland environment

It is important to remember, throughout the analysis of the dataset, that there was a broad division deriving from findings of field sampling analysis (Chapter 1). The elevated lochs is a broad term for lochs within which brown trout are infected by only one parasite;

Diplostomum baeri. The western lochs are lower lying marshland lochs within which brown trout sub-populations are infected with numerous parasite taxa; the cestodes *Diphyllobothrium dendriticum* and *Eubothrium crassum*, nematode; *Eustronglyides* spp., trematode; *Diplostomum baeri* and acanthocephalan; *Neoenchorhyncus* spp.

First analysis describes diversity indices of *Satr*-DAB gene within each subpopulation (Table 9). The subpopulation presenting the highest overall diversity is Lochan Fada (LFD; K:18.133). It also possesses the highest number of segregating sites (S; 46). The loch with both the lowest diversity observed across *Satr*-DAB (K: 5.2176) and the lowest number of segregating sites in pairwise comparison of subpopulation sequences (S: 17) is Loch na h-Oidche (LNO. Lochs falling within the elevated single species infected sub-populations had the following diversity indices Loch na Feithe Mugaig (LFM) (K: 6.94545, S: 18 and Pi:0.046), Lochan Fada (LFD) (K: 18.133, S: 46 and Pi 0.12001), Lochan Dubh nam Biast (LDB) (K: 9.96429, S: 28, Pi:0.65009) and Loch Airigh a'Phuil(LAP) (K: 6.55789, S:29 and Pi: 0.04343). The subpopulations that fell into the category of multiple species infection (western lochs) had the following observed diversity indices Loch Feur (LFE) (K: 12.01606, S: 36 and Pi: 0.07266), Lochan nam Breac (LNB) (K: 14.0757, S: 42 and Pi: 0.09322), Loch Coire nah-Airigh (LCA) (K: 9.67949, S: 30 and Pi: 0.0641) and Loch Gharbe Doire (LGD) (K: 10.1758, S: 30 and Pi: 0.06739). Sub-populations with multiple species infections demonstrated higher average diversity within population (K; 11.4867) compared to elevated single species infected lochs (K; 10.4001). Also the western loch sub-populations displayed a larger amount of segregating sites between alignments of *Satr*-DAB (S: 34.5) compared to single species infected lochs (S: 30.25).

Site	Length	S	hap	Hd	K	Pi
LFD	243	46	9	0.97778	18.133	0.12009
LDB	243	28	7	0.96429	9.96429	0.06599
LNU	243	33	5	1	15.7	0.10387
TWA	243	23	7	1	8.52381	0.05645
LMD	243	28	9	1	9.55556	0.06328
AAG	243	24	9	0.96364	7.8	0.05166
LCA	243	30	9	0.97165	9.67949	0.0641
FES	243	23	6	1	10.6	0.0702
LAP	243	29	17	0.98421	6.55789	0.04343
LGD	243	30	11	0.95604	10.17582	0.06739
LFE	243	36	12	0.96884	12.01606	0.07266
LNB	243	42	10	0.95455	14.07576	0.09322
LFM	243	18	6	0.8	6.94545	0.046
LNO	243	17	9	0.87179	5.21795	0.03456

Table 8; Diversity data associated with *Satr*-DAB gene across Scottish hill loch populations *n* –number of alleles, S- Number of segregating sites, Pi – nucleotide diversity, K – average number of nucleotide diversity per site, hap – number of haplotypes, hd– diversity between haplotypes observed.

The second summary of diversity indices from the loch system addressed the heterozygosity and number of alleles within individual subpopulations of the loch system. Table 10 describes the observed and expected heterozygosity of *Satr*-DAB gene within each subpopulation. Because the gene is inherited in a multi-allelic fashion, multiple alleles can be seen within F1 trout, these allelic counts at first can be isolated in original chromatographs of raw reads to give allele number, with further downstream analysis to show the observed and expected heterozygosity within sub-populations. Once again, data are divided into singular and multiple parasite infection lochs, this division of data allows

conclusions based on parasite infection diversity affecting heterozygosity and allele number within the sub-populations. Across all subpopulations, observed heterozygosity was greater than the expected heterozygosity. All subpopulations showed a greater number of alleles observed within the population than in samples, indicating that *Satr*-DAB exists in a heterozygous manner more than homozygous across all populations. Lochan Fada (LFD) indicated the greatest deviation from expected heterozygosity (0.39707) compared to observed (0.50001) (difference: +0.10294), with the loch indicating the least deviation Loch dubh na Biast (LDB) (difference: +0.01077). In terms of comparison between the two data sets, singular species-infected elevated loch subpopulations showed a greater average differential between observed and expected diversity (+0.188) compared to multiple species-infected western loch sub-populations (+0.04647) (Table 2).

Population	**Sample**	***N*a**	***He***	***Ho***
Western				
LFE	10	20	0.3971	0.5
LNB	10	20	0.2431	0.2539
LCA	12	24	0.3023	0.3581
LGD	10	16	0.5053	0.5855
Elevated				
LFD	10	20	0.3804	0.4287
LDB	10	18	0.3299	0.3689
LMD	12	22	0.3001	0.3678
LB	10	16	0.442	0.4729

Table 9; Summary statistics of genetic variability for sequenced MHC *Satr*-DAB locus within single and multiple species infected populations: N_a; number of alleles, H_e observed heterozygosity and H_e expected heterozygosity.

Results indicated differential heterozygosity characteristics and genetic diversity across the Gairloch system between the two designated grouping of subpopulations.

Phylogenetic analysis was performed to investigate lineage sorting according to geographical separation according to subpopulations and differential parasitism. Overall the final tree (Figure 17) construction showed little to no association to either factor. Topology showed next to no geographic-specific clades, with the only distinct population lineage within Loch a'Mhadaidh (LMD). Clade topology was not related to parasite specific species infection. With no general lineage sorting according to the western loch sub-populations (LNB, LCA, LGD, LFE) undergoing multiple species infection and elevated loch populations (LMD, LFD, LDB, LAP) (Figure 17). Bayesian tree construction is supported through strong nodal posterior values.

Figure 19; A detailed phylogeny using Bayesian inference of the *Satr*-DAB sequence alignment (252bp, 210 sequences). Lineages of Scottish samples have freshwater bodies abbreviated; LAM: Loch Atrigh Mhic Criadh, LAP: Loch Airigh a'Phuil, LDA: Loch Doire na h-Airighe, LFM: Loch Feithe Mugaig, LNB: Lochan nam Breac, LNO: Loch nah-Oidche. LNU; Nursery Lochan; LAM; Loch Arigh Mi Criadh; LFD; Loch Fada, LGD; Loch Gharbe a' Doire, AAG; Alt a' Glinne, LFE; Loch Feur, LB; Loch Buinachein, LCA; Loch Coirena h-Arigh, LDB; Loch Dum na Biast,, TWA; Talladale river, FES; Flowerdale estuary, LMD; Loch a'Mhadaidh and LSG; Lochan Sgeireach

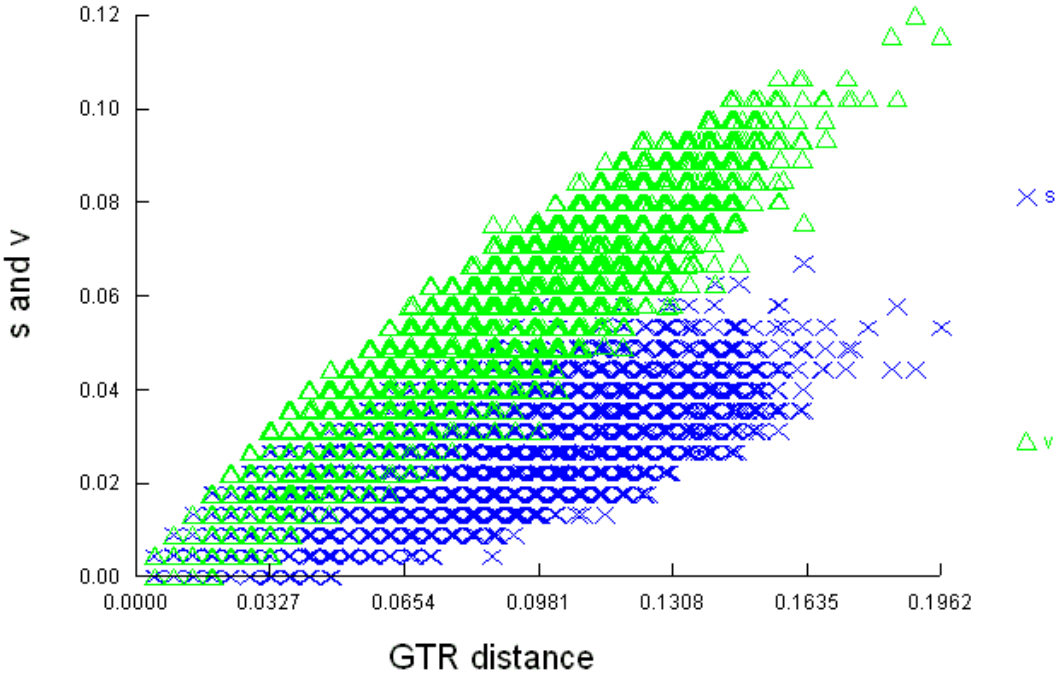


Figure 20: TV/TS graph according the GTR distance in correlation with number of transistions and transversions occurring per site across the gene alignment

From the initial diversity analysis of Satr-DAB, the gene presents as a highly polymorphic gene across populations, with an increased level of heterozygosity across the system. The overall pattern of diversity indices across the Gairloch system indicates an increased heterozygosity and genetic diversity occurring within the western loch populations, i.e. those infected with multiple parasite species. This sub-group displayed consistent increased average diversity per loch population (K) and an increased propensity for individuals to be heterozygous compared to elevated loch subpopulations. Although there is a deviation between diversity characteristics between the two sub-sets, this hasn't translated in phylogenetic cladal separation between sub-sets of populations or overall geographic lineage sorting across the complete dataset.

Resulting analysis from the DAMBE mutation saturation graph supported that the alignment did not present as mutationally saturated. This assumption can be made due to the positive linear result of the graph presented (Figure 20).

5.3.2 Divergence between Loch sub-populations in *Satr*-DAB

Across the Gairloch system, different levels of genetic diversity exist between subpopulations. Within systems of disparate sub-populations there is an expectation that through reproductive isolation, extended geographic isolation will correlate with a greater degree of genetic differentiation existing between sub populations of greater distance than those closer together. To determine whether this is present within the dataset a series of Mantel linear plots were used to analyze the role geographic separation plays in spatial genetic diversity between individual loch sub-populations. The first analysis addressed the potential genetic population structuring of *Satr*-DAB existing across the dataset in relation to geographic distance. This relationship was non-significant (p = 0.190 (>0.01), r^2= -0.092238) (Figure 18a). The second analysis was associated with the genetic diversity of *Satr*-DAB *versus* geographic distance between subpopulations. This was also non-significant (p = 0.724 (>0.01), r^2= -0.698122). As a consequence of these two analyses it may be surmised that there is no significant relationship between geographic isolation and genetic diversity seen in *Satr*-DAB (Figure 18b). The third regression analysis quantified the relationship between geographic distances *versus* the number of

nonsynonomous differences existing between subpopulations. In contrast this regression analysis showed a significant relationship between nonsynonomous mutations and geographic distancing in accordance to P value (0.0098 (<0.01) r^2= 0.678152), this indicates that with increased distance between subpopulations there is a significant increase in nonsynonomous mutation number between subpopulations (Figure 18c).

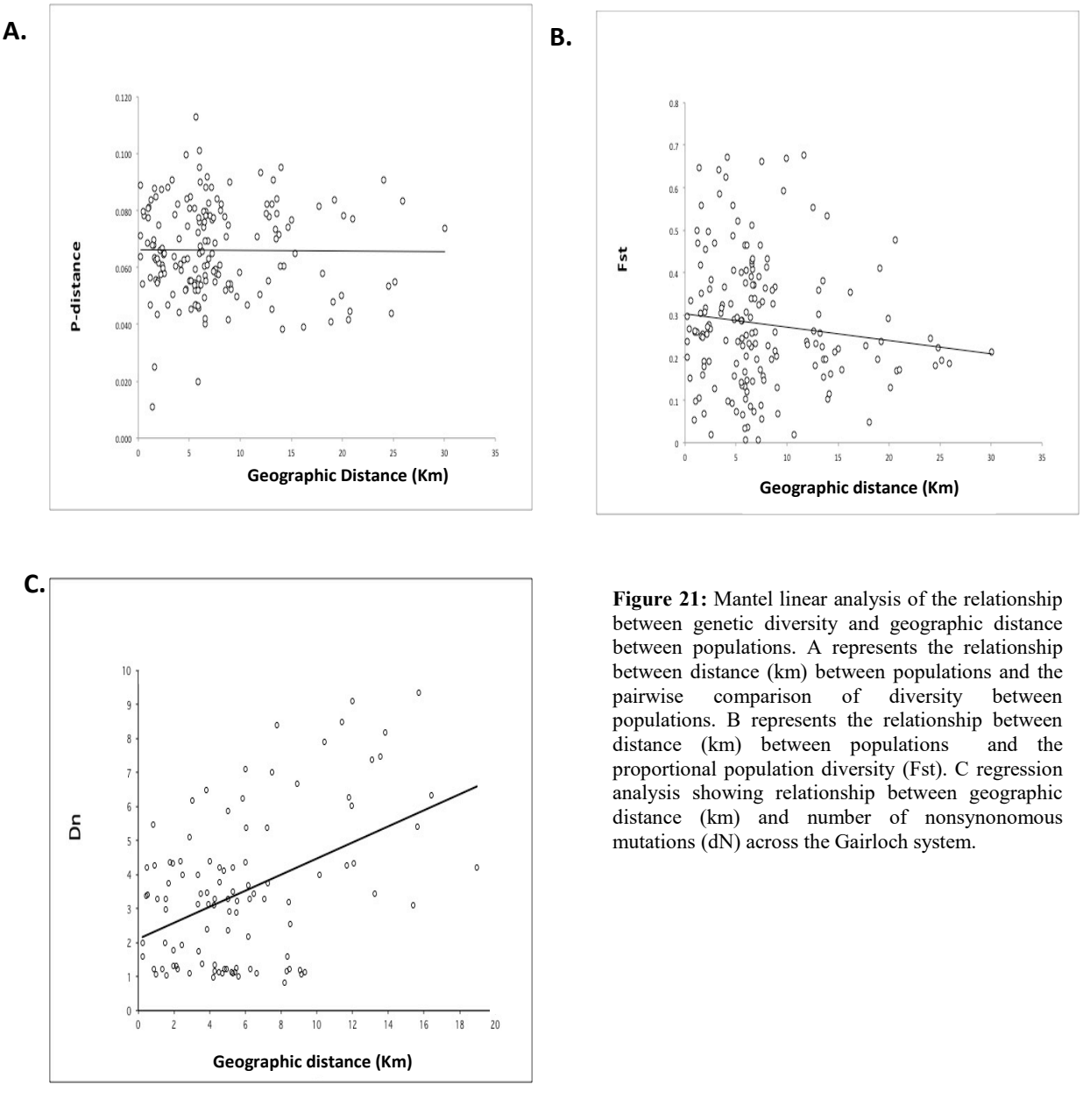


Figure 21: Mantel linear analysis of the relationship between genetic diversity and geographic distance between populations. A represents the relationship between distance (km) between populations and the pairwise comparison of diversity between populations. B represents the relationship between distance (km) between populations and the proportional population diversity (Fst). C regression analysis showing relationship between geographic distance (km) and number of nonsynonomous mutations (dN) across the Gairloch system.

5.3.3 Demographic history of Satr-DAB in S.trutta

As a gene, *Satr*-DAB has proven to be diverse throughout isolated populations of the Gairloch region. Definitive host movement throughout the loch system could provide one possible explanation for the maintenance of diversity of the gene across the population. With recent expansion events taking place within a locality, immune gene diversity can increase initially because new alleles enter a population creating an initial spike in genetic diversity followed by a gradual plateau as gene frequencies equilibrate under consequent change. Any deviation from this expected trend is shown as a near-bimodal distribution indicating no recent expansion in the population and is considered stable. To test this statistical significance of any deviation from expected distribution, statistical test *Tajima's D* and *Fu's F* is used to test neutrality of the gene. The third statistical test if a raggedness value; used to test deviation of observed and expected mismatch of diversity data.

Initial analyses were performed on the entire population with *Tajima's D* and *Fu's F.* These showed no significant deviation from neutrality. The alignment representing all sequences across Gairloch populations-presented *Tajima's D* was negative -1.32785 overall and significant $p = 0.041$ Fu's *F* calculations presented a similar pattern which was -3.8852 and a highly significant *p* value of $p= 0.011$, $p<0.001$. This value suggests that recent demographic population fluctuations such as expansion or retraction may have given rise to a large number of low-frequency haplotypes. The pairwise mismatch analysis across the entire dataset appeared bimodal in distribution, comparisons with the raggedness value; $r = 0.01326$, $P = 0.0001$ (Figure 19A) indicating a stable population. With different diversity indices between the elevated and western population subsets demographic movement analysis may go part way to explain potential differential diversity characteristics between the two species.

Western loch demographic analysis indicated no significant deviation from neutrality through *Tajimas's D* and *Fu's F* calculation. *Tajima's D* was negative overall -0.8252 with a *p*-value indicating a rejection of null hypothesis ($p<0.05$). *Tajima's D* negative value suggests an increased number of observed mismatched diversity compared to expected, indicating a recent expansion event, however, the negative *Tajima's D* value is not supported by its *p* value indicating a lack of significance in the result. *Fu*'s *F*

indicated no significant deviation from neutrality with an *F* value -12.529 and a rejection of null hypothesis ($p<0.05$). The pairwise mismatch analysis across the elevated lochs showed no recent expansion event occurring within the western loch populations with a bi-modal trending curve, $r = 0.02860$.

Elevated loch populations mirrored the lack of deviation from neutrality through *Tajima's D* (-1.28251) and *Fu's F* calculations (-20.108). Elevated loch data suggest a less demographic movement with similar bi-modal trending curve and a significant difference from expected and observed mismatch analysis ($r= 0.0207$).

Thus analyses of *Satr*-DAB do no support the model of host movement within subpopulations in either the elevated or western loch populations. As indicated in previous chapter (5.3.2) maintenance of diversity across the population may not lay within standing genetic diversity but the type of mutation occurring within the gene and the selection acting on the gene in subpopulations.

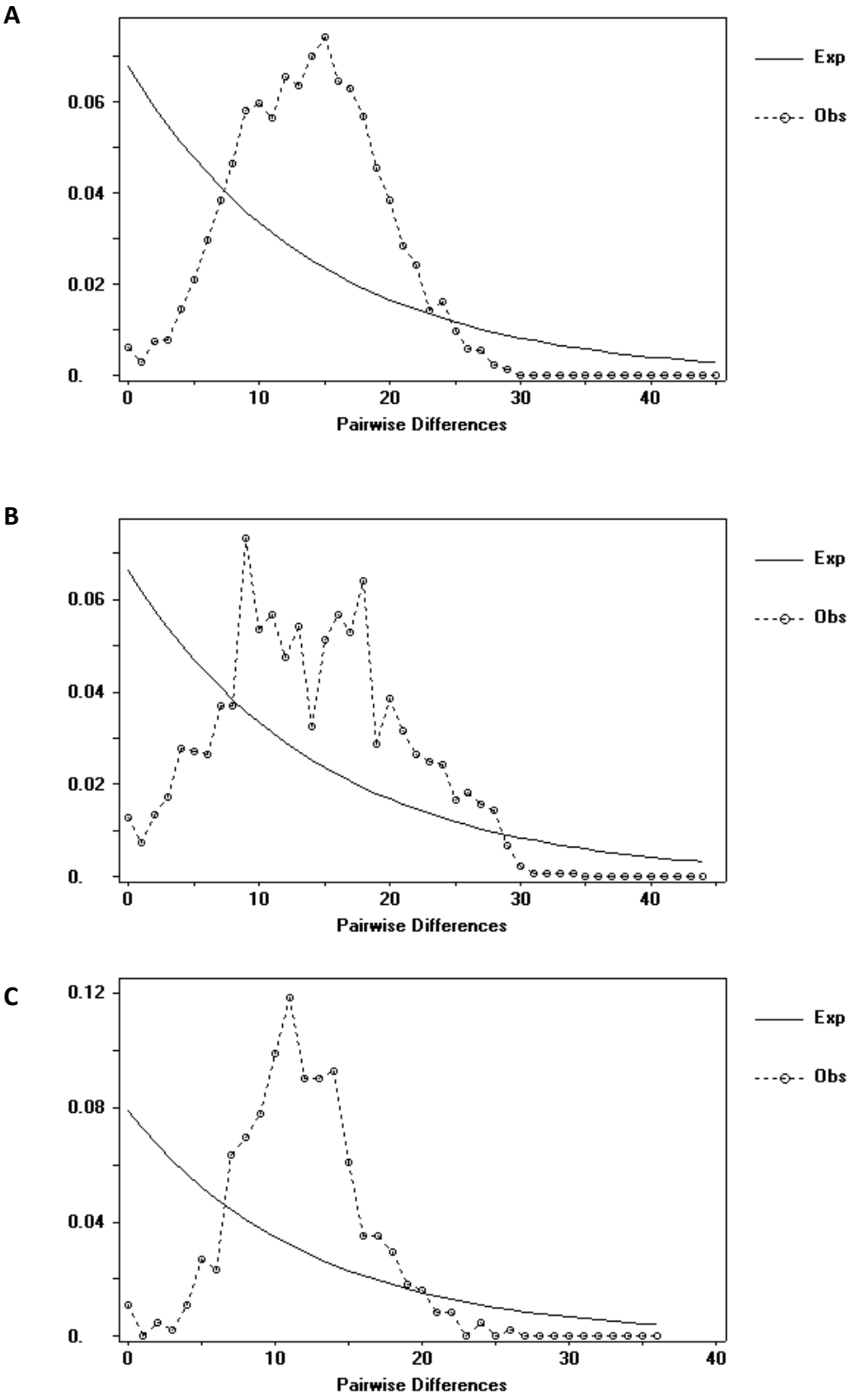


Figure 22; Mismatch analysis graphs to demonstrate historical demographic movement and potential population expansion events occurring across Gairloch population. An Entire population dataset, B elevated loch populations, C western loch population

5.3.4 Detection of selection in *Satr*-DAB

The diversity that is maintained in *Satr*-DAB genes is not associated with geographic distance between subpopulations or recent invasion/expansion events across the freshwater system. There is, however there is a clear relationship between diversity and number of nonsynonymous mutations between subpopulations.

MHC II as an immune gene is potentially under a great deal of selective pressure from pathogen insult, the presence of positive selection can be tested by testing the ratio of nonsynonomous mutations (dN) and synonymous mutations (dS). With dN mutations contributing to the eventual translated amino acid, the accumulation of dN mutations within a gene is denoted as being under "positive selection" with a dN/dS ratio >1, where the gene is under purifying selection the ratio is <1. Across the entire dataset in all sub-populations exhibit positive selection values (Table 3) with all dN/dS values for each subpopulations being >3.

Site	length	dN	dS	dN/dS
LFD	243	120.65	38.35	3.146
LDB	243	125.83	39.17	3.2124
LNU	243	124.43	37.57	3.3119
TWA	243	117.21	38.79	3.0216
LMD	243	117.13	35.87	3.2654
AAG	243	116.14	33.86	3.43
LCA	243	111.45	32.55	3.4239
FES	243	116.81	36.19	3.2276
LAP	243	113	34	3.3235
LGD	243	156.02	39.96	3.9044
LFE	243	159.85	39.07	4.054
LNB	243	158.22	38.56	4.1647
LFM	243	118.2	37.8	3.1269
LNO	243	118.79	37.21	3.1924
LB	243	112.88	32.2	3.4947

Table 10; Table displaying selective values across Gairloch subpopulations. Length; length of *Satr*-DAB gene, dN; number of nonsynonymous mutations across the subpopulation, dS; number of synonymous mutations across subpopulations and dN/dS; ratio of dN/dS total.

Across all populations *Satr*-DAB is undergoing positive selection with dN mutations being associated with variation in the final protein construction. It suggests that further investigations should shift emphasis onto the final protein sequence rather than gene sequence variation. The functional aspects of the ABS are crucial to ensure sufficient binding potential with parasite antigen. Positive selection not only acts on the entire gene but also can be examined on a codon-to-codon basis, investigating specific translated proteins for degree of positive selection acting on that protein. In order to test this the consensus *Satr*-DAB protein was constructed using likelihood amino construction of *Satr*-DAB genes for both differential subsets of Western and Elevated lochs. This was performed to analyze positive selection occurring on individual codons within populations experiencing differential parasitism. Within both the first 18 amino sites

showed no evidence of experiencing any form of selection according to the CODEML analysis; within section from 18 to 67 only 10 sites were experiencing no form of selection (27,28,31,33, 35,36,46,57,58,63,64,65). According to dN/dS ratios the elevated loch populations contained 6 codons under positive selection (21,23,24,34,52,67) compared to 10 under positive selection in western loch populations (22,23,29,32,37,40,41,47,50,51).

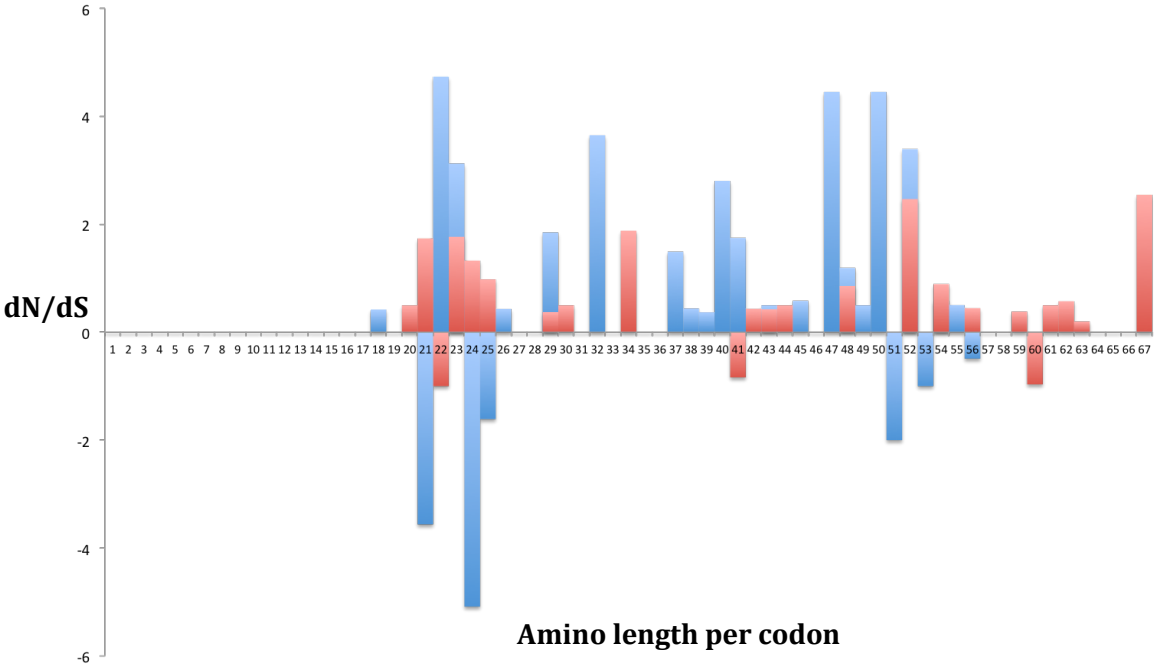


Figure 23; Codon-specific dN/dS ratio on the *MHC* II from PAML (CODEML) analyses. The dN/dS ratios are based on model M8 for all-haplotypes within divided dataset. Dataset was divided based upon populations infected with multiple species of parasite against populations infected with only singular parasite species. Colour coding denotes dataset western dataset highlighted in blue and elevated loch populations highlighted in red.

Across the two subsets selection is acting on the final translated amino sequence. The western lochs have demonstrated an increased numbers of codons under positive selection compared to the elevated loch populations. Three dimensional mapping of the ABS indicated increased level of selection acting upon codons translates to a variable amino sequence motif within the final translated protein. In this case, ABS 3-D design was mapped using a murine model (Wieczorek *et al*., 2017). Murine immune genes and MHC II structures have been subject to sufficient research to enable construction of such a model whilst those of salmonids have not. Even so, it was deemed that such a non taxon-specific model was still likely to be informative. A consensus protein 3-D was constructed using likelihood models and adjusted dN/dS values for each codon within consensus modelling. The elevated loch population demonstrates lower proportions of proteins under positive selection (supplementary table 1) compared to western loch consensus proteins (supplementary table 2). Additionally the increases presence of Lysine and Arginine across the ABS of western loch populations indicate an additional 3 hypothesized hydrogen bonds within the ABS binding site, indicating a region of increased binding potential within western loch subset (Figure 21).

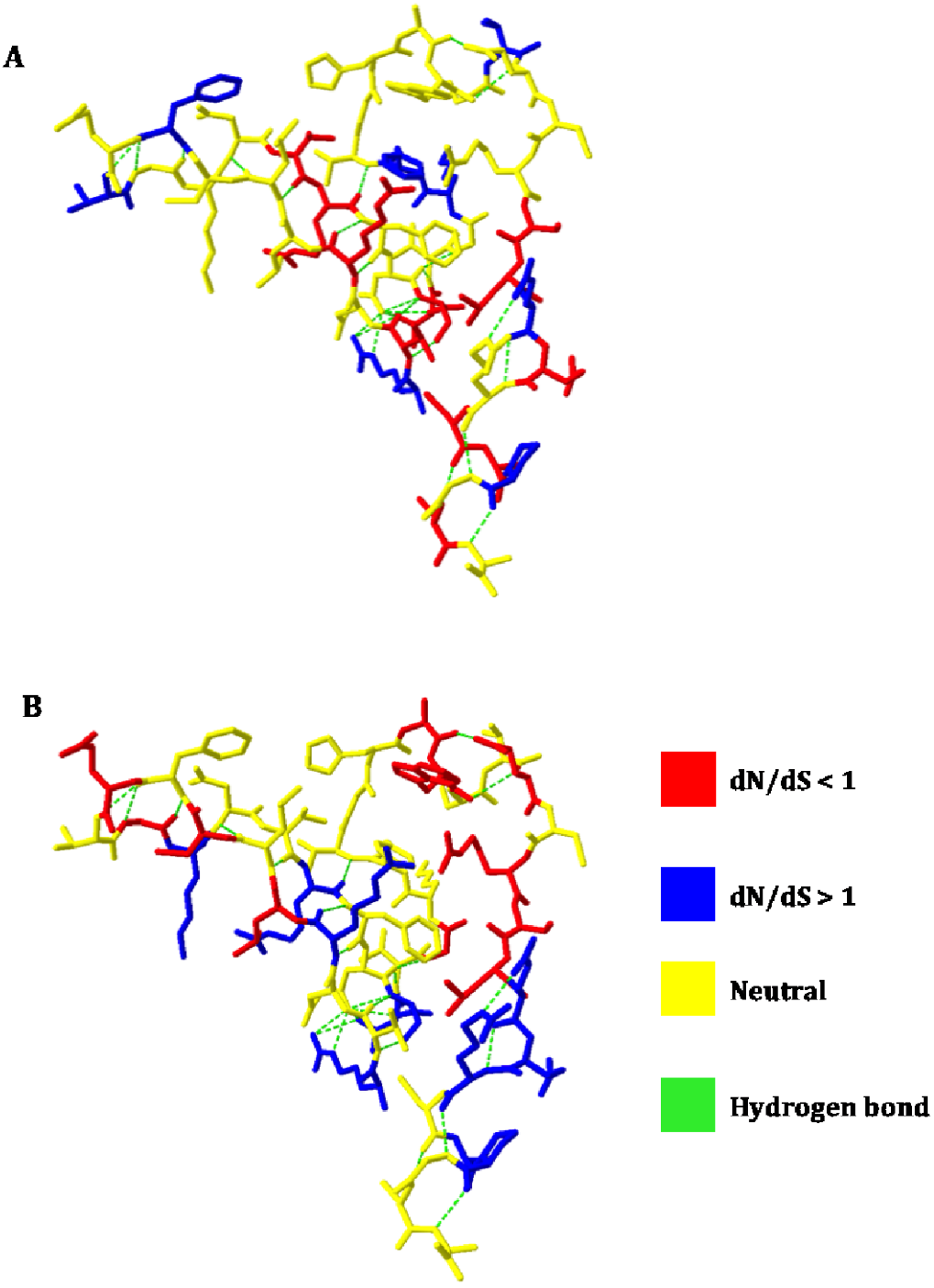


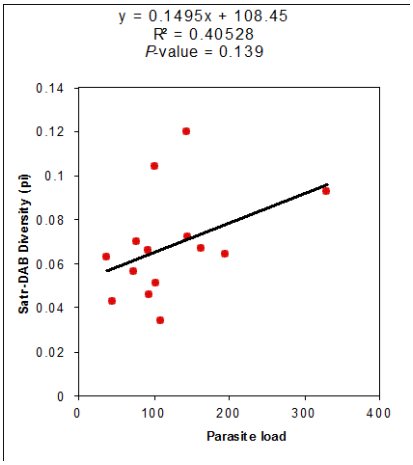
Figure 24; 3D protein structure translation of *Satr*-DAB MHC II exon 2. Colours denote dN/dS value outcome of protein codon; yellow; neutral, red; dN/dS<1, blue; dN/dS>1 and green; hydrogen bonds on protein. Structure denoted 'A' is consensus protein structure under singular parasite species infection pressure (elevated), structure denoted 'B' is consensus protein structure under multiple species infection presence (western).

5.3.5 Association between MHC variation and infection

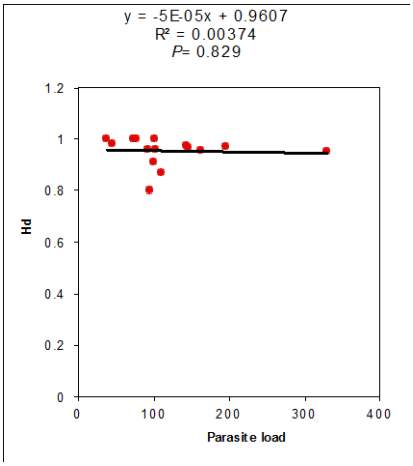
The comparative observations between the two subsets indicate differential characteristics in terms of effects of positive selection and ABS binding amino motif. As described in previous sections western loch populations are undergoing differential parasite infection compared to elevated loch populations (chapter 1). So far outcomes from *Satr*-DAB analysis reveal an increase in nonsynonomous mutations occurring within *Satr*-DAB in western loch populations, with these indicating an increased level of binding potential within ABS protein structure. Selection is acting in a differential manner within the two populations, with the only variable between the two being parasite infection diversity.

The regression analysis utilized parasite infection prevalence data within the loch system to analyze the relationship between parasite infection level with diversity indices and selective properties within *Satr*-DAB. Parasite load within host can exhibit direct correlation with individual host health and overall population health. The first regression plot (Figure 22A) shows relationship between *Satr*-DAB diversity and parasite load, with an overall lack of statistical relationship (P=0.139, >0.05).

A



B



C

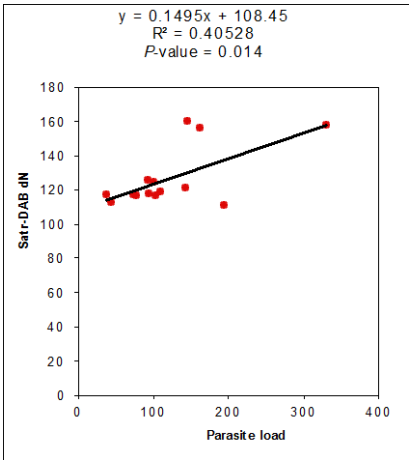
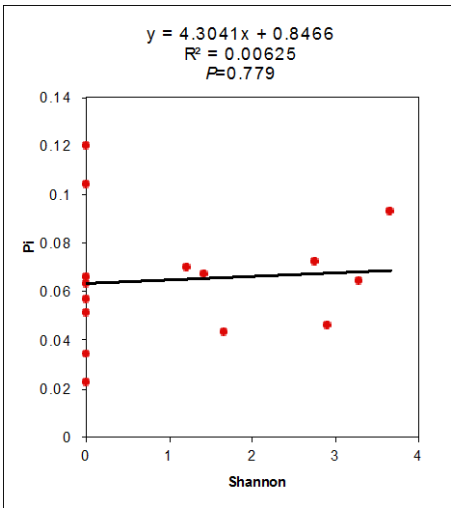


Figure 25; Regression analysis of *Satr*-DAB diversity indices in relation to parasite load in three separate regression analyses. A; *Satr*-DAB vs Parasite load, B; *Satr*-DAB Hd vs Parasite load, C; *Satr*-DAB dN vs Parasite load

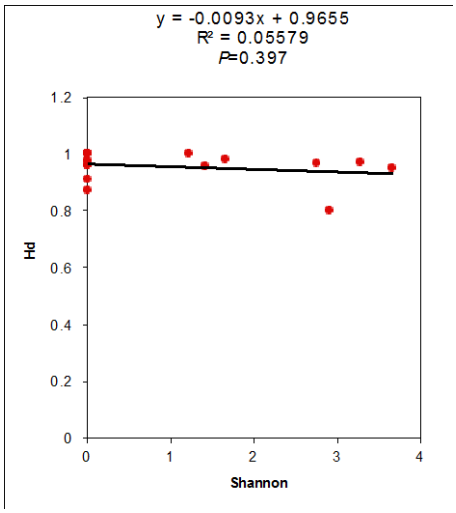
This is not true for the third regression analysis (Figure 22C) with nonsynonomous mutation level across subpopulations showing significant statistical relationship with *Satr*-DAB diversity (P= 0.014, P>0.05). *Satr*-DAB haplotype diversity (5B) shows no significant relationship with parasite load (P=0.829,)

The second set of analysis investigating parasite infection diversity investigates parasite diversity data within each subpopulation (chapter 1). Diversity values are placed upon subpopulations according to the diversity of parasites present within host populations within individual lochs. Diversity figures are calculated using Shannon diversity equations to denote diversity number per population. Shannon diversity values are used to compare the relationship between parasite infection diversity and *Satr*-DAB diversity, nonsynonomous mutation and *Satr*-DAB haplotype diversity. The first regression plot reveals no statistical relationship between *Satr*-DAB diversity (pi) and Shannon diversity values (P=0.779,) (Figure 23a), this is also the case with *Satr*-DAB Haplotype diversity (P=0.397,) (Figure 23b). However, a statistical relationship exists between the accumulations of nonsynonomous mutations within *Satr*-DAB in response to diversity of parasite infection (P=0.048,) (figure 23c), adding significance to previous selection analysis of ABS, with an increased number of codons under positive selection within western lochs undergoing a more diverse parasitic insult.

A



B



C

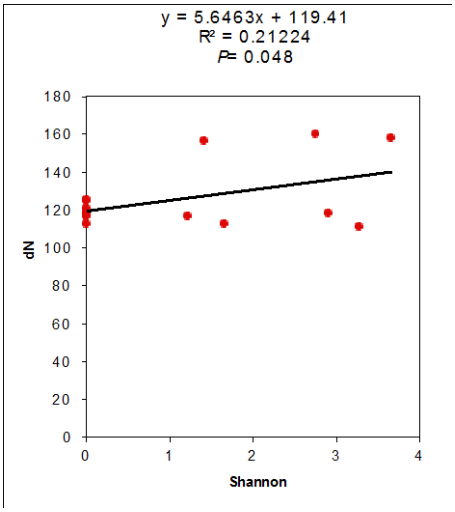
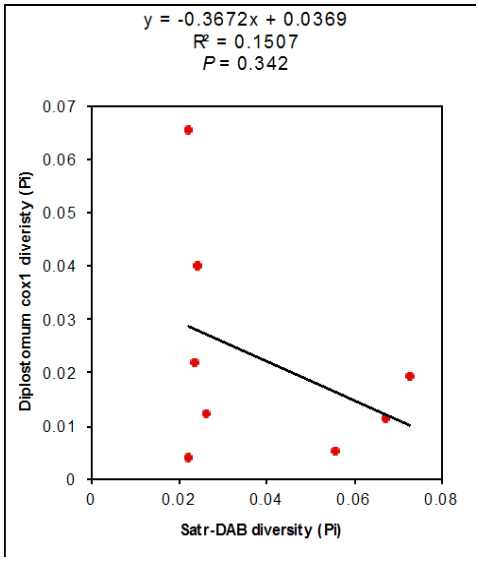


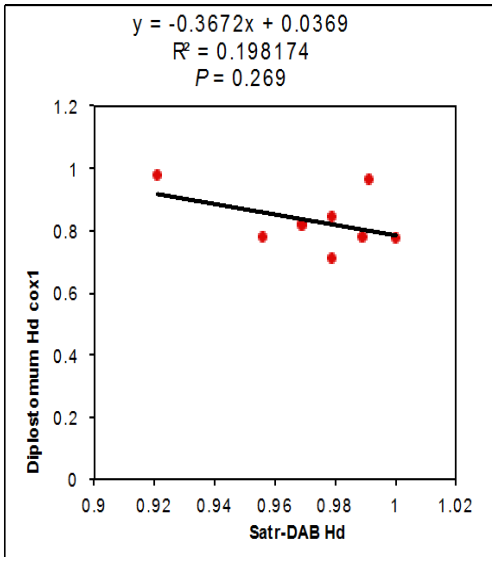
Figure 26; Regression plots to investigate the relationship between parasite infection diversity and *Satr*-DAB diversity (pi) A, haplotype diversity B and nonsynonomous mutation number C

The final set of analysis looking at the relationship of *Satr*-DAB and parasite infection investigates the effects of the most prevalent parasite within the system, *Diplostomum baeri*, upon MHC II diversity. The diversity indices of *Diplostomum* are taken from previous chapter (chapter 3) and use genetic diversity (pi), haplotype diversity (Hd) to investigate possible parasite-driven selection occurring from widespread *Diplostomum* infection. First regression plot shows no significant statistical relationship existing between *Satr*-DAB diversity and diversity of *Diplostomum*. spp across the entire dataset ($P = 0.342$, $P > 0.05$) (Figure 24a). There is also no significant relationship existing between the haplotype diversity of *Satr*-DAB and diversity of *Diplostomum* spp ($P = 0.269$, $P > 0.05$).

A



B



C

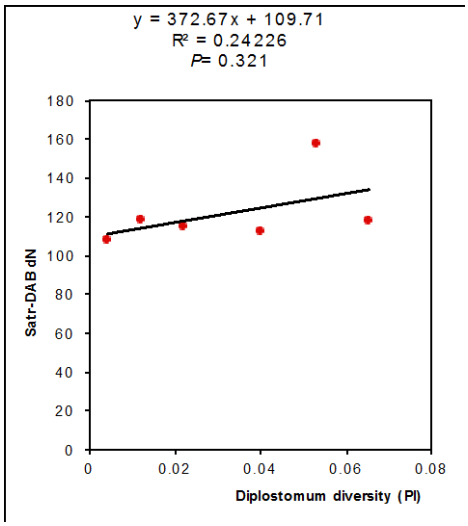


Figure 27; Regression plots investigation the relationship between A; *Diplostomum* spp genetic diversity (pi) and *Satr*-DAB diversity (pi) and B; *Diplostomum* spp haplotype diversity and *Satr*-DAB haplotype diversity. Genomic data deriving from *Diplostomum* all derive from *cox1* genetic markers. C *Diplosotmum* spp diversity vs *Satr*-DAB dN

5.4 Discussion

The investigation found signals of positive selection within the *Satr*-DAB gene across all sampled populations of brown trout with values associated with genetic diversity of *Satr*-DAB indicating clear inter-population differentiation. One constant throughout the Gairloch system was the strong signal of positive selection acting on the MHC *IIβ*, with all lochs showing dN/dS>3 on *Satr*-DAB gene. The dN/dS values were corrected for using p value; with normalized dN/dS mean being 3.3. It is reasonable to assume immune system genes would be under strong selective pressure. Positive selection indicated through elevated dN/dS values have been demonstrated in numerous other taxa (Radwan *et al*., 2014; Zueva *et al*., 2014;Wezner *et al*., 2016) these studies have emphasized the importance of parasites as selective influences acting upon the MHC gene complex.

Phylogenetic reconstruction revealed little to no geographic structuring of different populations across the region suggesting that although selective pressure is strong, it has not lead to significant differentiation in terms of topology. Many broad-scale studies have attempted to elucidate phylogenetic outcomes of salmonid species utilizing MHC II genes (Grimholt *et al*., 2003; Mjaaland *et al*., 2005; Kjøglum *et al*., 2006; Croisetière *et al*., 2008; Lamaze *et al*., 2014b; Ciborowski *et al*., 2017; Schenekar and Weiss, 2017). Multiple studies elucidated phylogenetic population structuring correlating with specific bacterial infection presence (Grimholt *et al*., 2003b; Kjøglum *et al*., 2006; Croisetière *et al*., 2008). Phylogenetic structuring was also observed when comparing stocked brood vs wild brood fish within freshwater systems (Mjaaland *et al*., 2005; Lamaze *et al*., 2014a; Schenekar and Weiss, 2017). In reference to diversity indices of *Satr*-DAB within the Gairloch hill loch system the gene presents as highly polymorphic, not providing sufficiently conserved alignments to elucidate phylogeography such as that found within brown trout inhabiting river catchments in north west Spain A study by Thomas and Turner, 2008, also observed population structuring of MHC II within 10 Arizona populations of endangered Gila trout (*Oncorhynus gilae gilae*). Wider phylogeographic investigation indicated global population structuring in salmonid MHC II including *Salvenius alpines* (Arctic charr) (Conejeros *et al*., 2008) and *Salmo salar* (Atlantic

salmon) (Langefors *et al*., 2001). Within the Gairloch system this has not occurred, with trout populations not showing evidence of long-term stability of MHC II haplotype. Although with signals of positive selection exist across Gairloch subpopulations, it is functional adaptation or nonsynonomous mutation that promotes differentiation rather than traditional lineage sorting across populations.

The wild brown trout populations within the study are highly fragmented and found within isolated locations with wholly different environmental characteristics in respect to hydrology, fauna and potential pathogenic infection. Gairloch system brown trout need to adapt to differential parasitism exhibited between hill loch populations via differential MHCII allele frequencies. Such differentiation, therefore, may be considered evidence of differential fitness advantage in response to varying parasitic milieu; i.e. increased parasitic selection pressure positively associated with increased heterozygosity. The current study demonstrated all sites exhibited positive selection acting upon MHC antigen binding sites, this indicates that selection rather than recombination exhibits a much stronger force in maintaining the diversity of MHC II. This result concurs with data from German and Irish populations of wild brown trout whose MHC I locus demonstrated higher allelic heterozygosity compared to farmed conspecifics (O'Farrell *et al*., 2013; Schenekar and Weiss, 2017).

The genetic structure of populations is the product of gene flow, genetic drift, selection and mutation. The analysis of ecologically important genes such as MHC II holds great promise for studying natural selection and local adaptation. The genetic structures of riverine populations have frequently been affected by genetic flow (Hindar et al., 2004). Studies have recently given empirical evidence that salmonid population gene flow present in a asymmetric manner, moving from larger populations in a system to small populations within the natural setting (Hansen et al., 2002; Manier and Arnold, 2005) This form of gene flow frequently structures salmonids into differentiated populations (Palstra et al., 2007; Gomez-Uchida et al., 2009) but this feature does not exist within the Gairloch system in relation to *Satr*-DAB. On the genomic level adaptive divergence between populations is not present, with a lack of geographic topology existing across the

Gairloch region. Genetic drift does not play a role in the evolution of *Satr*-DAB within the Gairloch populations; this is in stark contrast to previous findings of brown trout populations in northern Spain. Campos et al (2006) performed a study into the population structure of brown trout within nine isolated riverine populations within single river drainage; genetic variation occurred in correlations between MHC diversity and microsatellite markers throughout populations in the system suggesting that genetic MHC variation observed seems to have been shaped by neutral forces such as genetic drift, rather than selection. However, the study does not take into account differential factors affecting the MHC diversity. Because all populations studied within the northern spainish rivers were infected by only singular species of pathogen, neutral forces may have had a greater impact on MHC diversity compared to adaptive measures against a diverse pathogen milieu, such as Gairloch trout populations.

Tests at the genomic and protein level indicated the presence of strong positive selection within Gairloch trout populations. These signals were more pronounced within western populations, which were infected by multiple parasites species and thus possessed a larger proportion of codons under positive selection. This pattern of increased selection within the ABS can be interpreted as a result of pathogen-driven selection leading to high levels of intra-population diversity associated with increased diversity of pathogenic insult. The individual codon analysis supported a number of codon positions under strong levels of selection compared to other codons. Even though parasite-driven selection of MHC alleles has been demonstrated in numerous 'hatch stock' vs. wild populations in brown trout (Schenekar and Weiss, 2017), sea trout (Byrne *et al*., 2002), Gila trout (Peters and Turner, 2008) and Atlantic salmon (Mjaaland *et al*., 2005), wherein there is consistent increase in heterozygosity in wild populations compared to stocked/farmed populations (Byrne *et al*., 2002; Lamaze *et al*., 2014a; Mjaaland *et al*., 2005; Taylor, 1991), this is the first study to elucidate differential parasite-driven selection acting on MHC II within wild brown trout populations. Importantly, these results indicate that such farmed vs. wild comparisons do reflect operation of selection processes occurring between wild populations rather than purely reflecting processes resulting from the

artificiality of aquacultural systems. This has implications for the acquisition and maintenance of salmonids for aquaculture and for salmonid population conservation.

One constant throughout the system was the high level of *Diplostomum* infection. Typically, genetic diversity of a highly prevalent parasite can contribute significant selective pressure on host MHC allelic diversity (Kalbe *et al*., 2006; Fraser & Neff, 2009). A selective sweep of an allele conferring greater infective success across populations of *Diplostomum* spp. would lead to an effective host immune response, thus maintaining adaptive MHC diversity. Although not previously tested within salmonid species, selective pressure exerted by *Diplostomum* spp. can initiate a hereditary immune response in *Gasterosteus aculeatus* (three-spine stickleback) (Kalbe and Kutz, 2006). Within the Gairloch system no statistical relationship has been proved between *Diplostomum* diversity and *Satr*-DAB diversity. Although results indicate no statistical significance it does not discount the biological significance that parasite infection presence exerts on MHC maintenance.

5.5 Conclusion

Data acquired within this study suggest adaptive immunogenic traits of wild brown trout allow isolated populations to habituate to pathogen presence within their immediate environment, with a direct selection bias to differential infection presence between two isolated groups of fish undergoing diverse parasitic insult. These results are the first of their kind and have implications, not only for the trans placement of wild populations of brown trout but would also suggest that exposure of non-native stocked brown trout to novel pathogens is inadvisable when trying to boost numbers of wild resident salmonids. Understanding the way salmonids adapt to minimize potential adverse fitness affects posed via parasite infection should also be considered with both the potential stocking procedures of non-native fish and ensuring sustainable practice in aquaculture.

Isolated populations of brown trout utilize immunogenic adaptive traits to co-exist with long-term parasite insult, with a diversity of infective potential acting to maintain crucial genetic diversity within host MHC genes. Even with MHC genes adapting to parasite diversity, infective presence especially *Diplostomum baeri*, still persists within host populations, the continued infective success of *D. baeri* indicates its continued successful evasion of the host immune system. The molecular basis for immune evasion in freshwater trematodes is predominantly due to intracellular antigens down-regulating immune response mechanisms (Maizels, 2009). Genomic techniques have highlighted said antigens within multiple trematode species (Hemler, 2003; Levy and Shoham, 2005b; Chalmers and Hoffmann, 2012), however the presence of such immunomodulatory antigens have not yet been confirmed in *D. baeri*. Intracellular antigen presence within *D. baeri* could, thus, be a factor in the species' high prevalence within Gairloch brown trout populations.

6 The construction and annotation of the Mitochondrial genome of *Diplostomum baeri*

6.1 Introduction

6.1.1 Parasite genomics

The advent of genomic-based studies of parasites has allowed expansion of knowledge associated with interactions between parasite and host (Tarleton and Kissinger, 2014). Basic PCR-based analysis has allowed identification of parasite species through genetic markers and functional genes throughout host populations within endemic areas. A relatively recent form of genomic technique has been Next-generation sequencing; its introduction has now provided multiple ways to utilize parasite genomics and large-scale biological data to give insights into comparative transcriptomics and proteonomics between populations (Preidis and Hotez, 2015). Even though parasite genomics now provides a fast and accurate method to investigate parasitic disease, the majority of parasite genomic information has been obtained from human-infecting parasites. The open source genetic databank "genbank" (https://www.ncbi.nlm.nih.gov/nuccore) holds over 172,046 annotated sequences and chromosomes for the causative agent of African sleeping sickness; *Trypanosoma cruzi* and 159,115 for Schistosomiasis causing; *Schistosoma japonicum*, however the highly pathogenic fish ectoparasite *Gyrodactylus salaris* is only represented by 2282 annotated genes. With the continued reliance upon aquaculture and fisheries to supply food, jobs and economic input to communities there is a clear need to ensure primary fish stocks are kept healthy (Piasecki *et al*., 2004; Whitmarsh and Wattage, 2006). The expanded use of NGS technologies on freshwater parasites would increase knowledge regarding species delineation but also aid in highlighting potential vaccine and drug targets as seen in the case of *Dirofilariaimmitis* (heartworm) (Godel *et al*., 2012) and *Fasciola hepatica* (Dalton *et al*., 2003). The increased knowledge of genomics associated with aquatic parasite infection dynamics could aid in the maintenance of farmed fish health, via improving long-term fish immunity rather than an over-reliance upon pharmacological intervention. It would not only be the aquaculture industry that would benefit from increased aquatic parasite genomics research, conservation efforts could become significantly more efficient when understanding potential parasitic diseases affecting high value stocked fish and

preserving endangered fish species.

6.1.2 Mitogenomics and molecular epidemiology

Since the initial studies of parasite molecular epidemiology the mitochondrial genetic marker cytochrome oxidase 1 (*cox1*) has been used to aid in delineating between species of closely- related parasites (Moszczynska *et al*., 2009b). In an effort to increase the accuracy of species identification between closely related species research has utilized many forms of genetic markers including repeat motifs of micro satellite markers (Criscione and Blouin, 2005; Keeney *et al*., 2006) and nuclear markers (Georgieva *et al*., 2013c). In recent years another potential sources of molecular markers in parasites has been mt genomes. Despite a multitude of potential mitochondrial genes being isolated as potential markers (Roy, 2014) few authors have considered alternative mt sequences for studies, with the majority of molecular investigations of parasites relying on the *cox1* barcoding region. A comparative mitogenomic approach could identify potential genetic targets which would have highly variable characteristics needed to investigate species identification in closely related subpopulations. With many molecular-based studies (including chapter 2) (Galazzo et al., 2002; Selbach et al., 2015a), revealing varied results from using varied genetic markers, the potential of utilizing mt genomes for the discovery of novel and potentially more accurate markers should present a well-recognized prospect. The structural content mitochondrial genomes of trematodes include 22 transfer RNA's, 12 protein coding genes, to ribosomal genes and vary between 13-17,000 bp in length (Brabec et al., 2015b; Littlewood et al., 2006; Ma et al., 2016). Current evolutionary studies between species have outlined the arrangement of genes and tRNA throughout the length of the mt genome which has been a significant factor in delineating between species (Telford *et al*., 2000). As with many parasite genomic studies, the majority of parasite mt genomes have been isolated within medically-relevant parasites (T. H. Le *et al*., 2000; Littlewood *et al*., 2006; Ma *et al*., 2016) and agricultural (Shekhovtsov *et al*., 2010) but not from freshwater parasites. *Diplostomum* represents an interesting model genus for population genetics analysis due to its species' wide distribution and complex asexual and sexual life cycles (Georgieva *et al*., 2013). To overcome potential hurdles in *Diplostomum* species delineation there is a need to

supplement and analyze new potential markers to aid in molecular identification. Characterizations of mitochondrial genomes have advanced identification techniques of many other medically important trematode species (Brabec *et al*., 2015; Le *et al*., 2000; Littlewood *et al*., 2006; Shekhovtsov *et al*., 2010), with advances in NGS sequencing technologies allowing for the acquisition high quality of genomic data from which mitochondrial genomes can be readily identified for further downstream analysis. One example where such analyses have occurred to aid in the accurate identification of *Diplostomum* species, mitochondrial genomes have been constructed of two common species *D. spathaceum* and *D. pseudospathaceum* with tissue samples from adult worms infecting definitive host; *Larus ridibundus* (black-headed gull) (Brabec *et al*., 2015), however with most pathogenic impact exerted via *Diplostomum* infection within host intermediate fish populations the data acquired from metacercarial genomic studies would further advance understanding of the crucial intermediate stages of *Diplostomum* species.

6.1.3 Aims and objectives

The genomic investigation of *Diplostomum* was performed in a two-part study. The aim of the first part of the study was to construct and analyze mitochondrial genome of *Diplostomum baeri*. This was followed by the development and assessment of molecular markers for epidemiological studies using mitochondrial markers. The objectives were:

- Construction and annotation of protein coding genes, tRNA ribosomal genes of *Diplostomum baeri* mt genome
- Comparison of mt genome across *Diplostomum* species for size, overall content and G-C richness and gene order
- Potential efficacy for mt genome to advance species delineation within *Diplostomum*

6.2 Methods

6.2.1 DNA extraction

The metacercarial stages of *Diplostomum baeri* were acquired from the optical lens of *Salmo trutta* (brown trout) from Loch Fada, Scotland (Lat; 57.6755201, Long; -5.2926635). DNA was extracted from 10 metacercariae of *D. baeri* using a DNeasy blood and tissue extraction kit (Qiagen). Samples were pooled to ensure adequate concentration of DNA was available and ensure downstream library preparation and further NGS procedures were performed using the highest quality and quantity of *Diplostomum* DNA. Elution of total DNA was isolated and quality checked using NanodropTM 2000 spectrophotometer, yielding 1.8ng/ul of total DNA in final DNA elution.

6.2.2 Whole genome sequencing

Whole genome sequencing was carried out using Illumina MiSeq at the Natural History Museum, London. Initial library preparation was performed using TruSeq Nano DNA Preparation Kit and run using MiSeq Illumina sequencer, amplifying 250bp long paired-end reads. The MiSeq next-generation sequencing technology uses fluorescent terminator dyes to detect bases during sequencing. Pooled sample was sequenced 4 separate times to improve overall genome coverage and data was combined into a single data set.

6.2.3 Sequence quality control and trimming

Initial raw read data was handled using Qiagen CLC genomics workbench to clean, trim and analyze data generated. The failed and low quality reads were removed based on quality filters and discarded with quality scores retained in the output. The quality scores associated with read quality on CLC genomics workbench are denoted using 'phred scale' this is a calculation based on error probability across the raw read dataset to inform the user of the based that have been trimmed. Sequences are timed on quality (=0.01), ambiguities (0) and length (<50). Trimming parameters were selected through scaling

algorithms to identify the most appropriate maximum coverage parameters whilst maintain the quality of final genome construction. Sequences over 50bp within read dataset are more likely to be unique within the genome, the 'uniqueness' of these sequences and denoted as K-mer ratio (Schatz *et al*., 2010). After trimming a quality report of the data is constructed for the user to examine overall quality of construction. During MiSeq procedures adapters are attached to cleave DNA inserts that are being sequenced. If DNA is shorter than the defined prearranged sequence length then universal adapters are attached to aid in the sequencing of the whole genome, these adapters must be trimmed during the quality control phase.

Adapters used:

TruSeq Universal Adapter;
5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT
CCGATCT
Truseq Adapter, Index;
5'GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCC
GTCTTCTGCTTG
Sequences were trimmed of adapters with alignment scores of; mismatch = 2 and gap cost = 1. These match scores provided most accurate results when checking discarded sequences once adapters were removed from the final dataset.

6.2.4 De novo mapping of reads

The de novo mapping of the *Diplostomum baeri* genome was performed using CLC genomics workbench. The assembly of reads into final de novo genome was performed using de Bruijn assembly algorithm graphs (Zerbino *et al*., 2009). The paired distance estimation is enabled to allow the mapping of paired reads at first into contigs sequence and then resolving larger repeats into scaffolds. Final summary of de novo efficacy is presented within nucleotide distribution analysis and contig length distribution within final mapping report.

6.2.5 Mitochondrial genome assembly and annotation

Downstream annotation to provided information on positioning of mt protein and tRNA-coding regions were performed utilizing a multitude bioinformatics software analysis programs. Initial mt protein-coding gene placement was revealed via the use of MITOS (http://mitos.bioinf.uni-leipzig.de/index.pyt) online genome annotation tool, tRNA genes were annotated onto *D. baeri* using tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/) online web server along with ARWEN tRNA gene prediction tool (Laslett and Canbäck, 2008). Placement of annotations was performed using Geneious MAFFT alignment plugin (Katoh and Standley, 2013) along with verified published genomes *Schistosoma japonicum* (NC_002544) and *Fasciola hepatica* (NC_002546). Alignment of mitochondria within the Diplostomidae was also performed using MAFFT alignment algorithm, along with the whole length of mitochondrial genomes of *D.spathaceum, D.pseudospathaceum* and *D. baeri*.

6.2.6 Mitochondrial genome phylogenetic analysis

Individual annotated genes of the mt genome were translated using Geneious v8.5 (Kearse *et al*., 2012) and resulting concatenated amino acid sequences of *D. baeri* were aligned manually using published genomes of Digenea. These included Diplostomidae; *D. spathaceum* (KR269763) and *D. pseudospathaceum* (KR269764), Schistomidae; *S. Japonicum* (NC002544), *S. spindale* (DQ157223) and *S. haematobium* (DQ157222) Echinostomatidae; *Hypoderaeum* sp. (KM111525), Fasciolidae; *F. hepatica* (NC002546), *F. gigantica* (KF543342). The cestode *Diphyllobothrium* was included as an outgroup.
All alignments created using MAFFT 7.122, with poorly aligned sites with extensively divergent regions being eliminated using Gblocks Server v.0.91b (http://molevol.cmima.csic.es/castresana/Gblocks_server.html). The tree construction was performed using Bayesian inference in Beast software v2.46, the evolutionary model used for analysis was GTR+I+G selected via bModelTest (Bouckaert, Heled and Kuhnert, 2014). Markov chain parameters were run for 10,000,000 MCMC generation's

sampled trees every 1,000 generations with 25% discarded as burn-in. Final Phylogenetic tree visualization was performed using FigTree v1.42 .

6.3 Results

6.3.1 Whole genome sequencing

The total number of paired end sequences was 19,829,002 before QC and after trimming produced 6,824,680. The total number of nucleotide sequences was 2,003,995,037 before QC and 1,734,130,204 after trimming and quality control sequence removal measures.
The genome size was estimated at 379-384 Mb by the program GenomeScope (Vurture *et al.,* 2017). The program fits a model of negative binomial distributions to the k-mer profile. The sequence data was assembled into 4,346 scaffolds with an N50 of 198kb (table 12).

6.3.2 Sequence quality control and trimming

Different quality score values were assessed across sequence and per base within the genome to maximize the quality of the final data set. The length distribution within the *de novo* construction describes the distribution of sequence length for the genome; the *Diplostomum baeri* sequence construction sequence lengths indicated lengths ranging from 50-300bp. After trimming the genomes retained no sequences below the length of 50bp (Figure 25). For each genome the G-C content remained similar post trimming with final content being around 38%. The G-C content per base indicated fluctuations across the reads within the first 80 bases, after trimming the G-C per base value improved in consistency, with an increase in G-C content in the first 10bp along with an increase per base score within 250-280bp across reads. The analysis of quality distribution across the genomes ranged from 14 to 39 PHRED score before trimming, after trimming and QC procedures this score rose to 25 to 39. The contributions of nucleotides to complete genome indicated a large fluctuation before and after trimming within the first 80 bases of each sequence, after trimming consistency of bases within beginning of each sequences increased with dips in guanine and adenine within latter portions of each sequence across read dataset. The over representation analyses of enriched primers calculates the ratios of observed and expected penta-nucleotides in the genome, this was

represented as an over representation of AAAAA within the 55-75bp portion of initial datasets, mainly owing to adapter sequences and sequencing errors, these decreased from 19% to below 1% after trimming and QC measures.

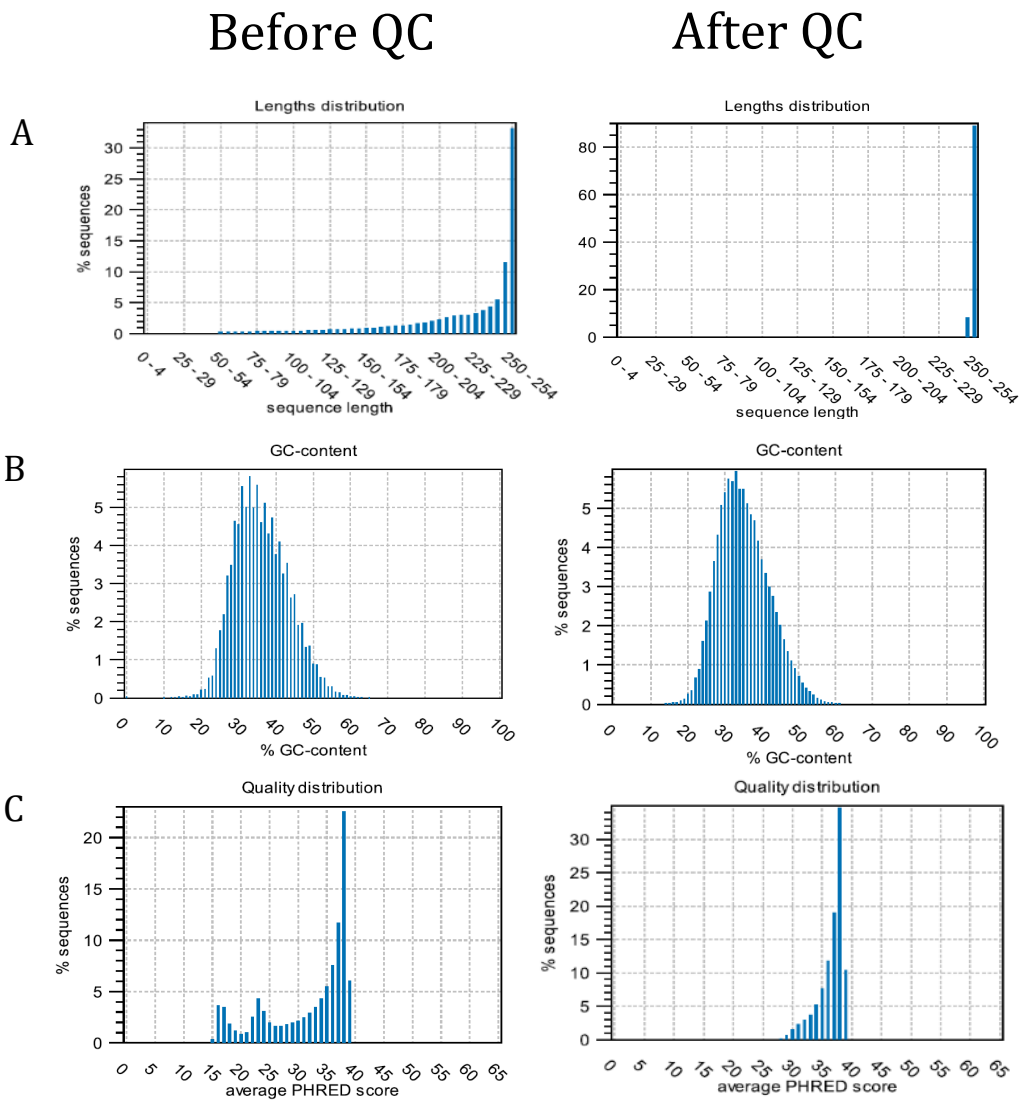


Figure continued …

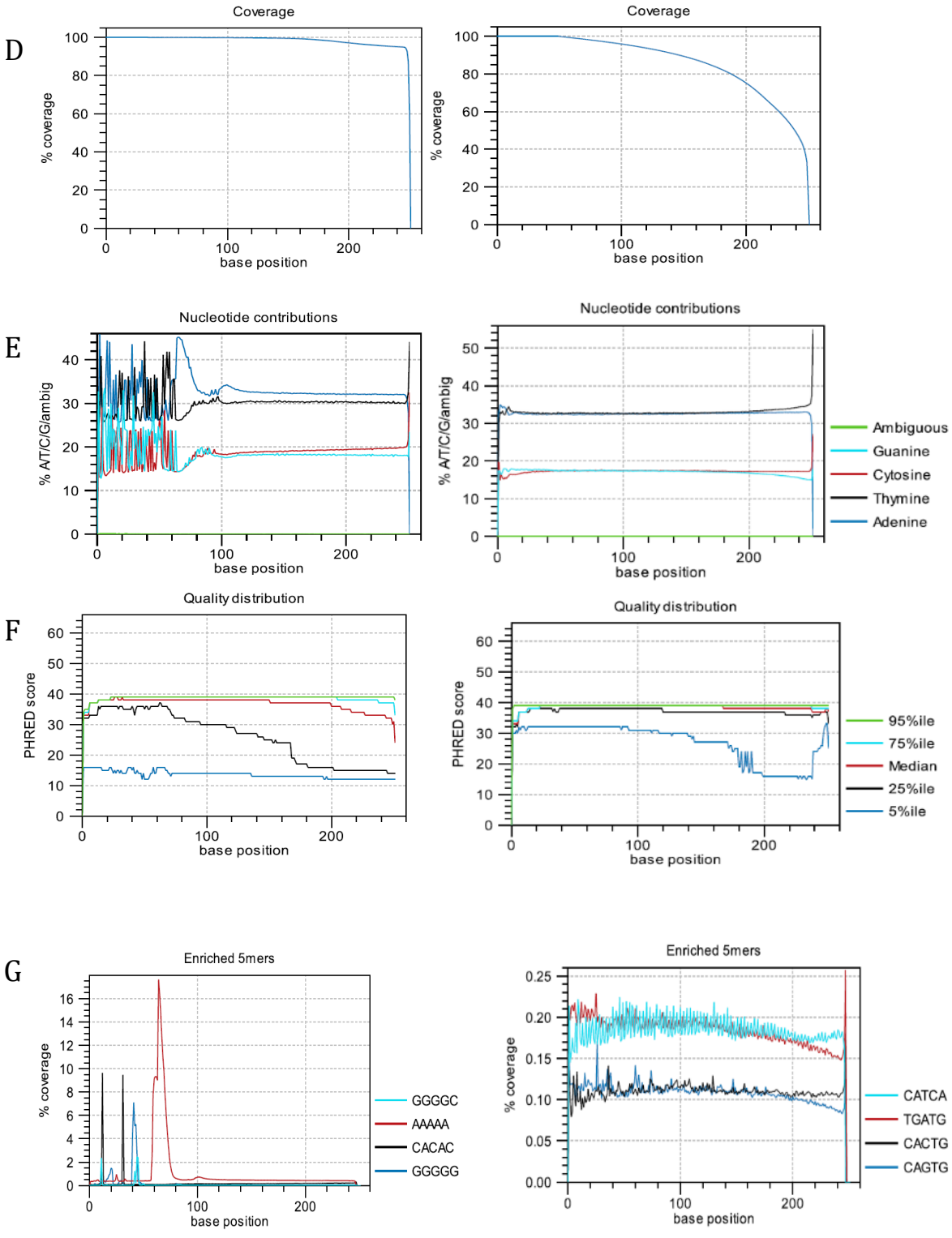


Figure 27; Resulting graphics from CLC genomics quality control workflow, graphs are divided between before and after trimming procedures and QC measures. The graphs are divided into; A: distribution of sequence lengths, B: GC content across sequences, C: PHRED average score across sequences, D: per base coverage %, E: Nucleotide contributions per base, F: per base PHRED score distribution, G: enriched 5mers across sequences.

Assembly criterion	*D. baeri*
Estimated genome size (Mb)	379-384
GC content (%)	33.6
Repeat rate (%)	35.11
Contig stats	
Total base pairs	358.3
Number of contigs	19 759
N50 length (kb)	108
Length of largest contig (kb)	188.3
Mean contig size (kb)	15.7
Scaffold statistics	
Number of scaffolds	4 111
N50 length (kb)	197.5
Largest scaffold (Mb)	339.1
Number > 1kb (% of assembly)	4,222 (100%)
Number > 10kb (% of assembly)	3,799 (99%)
Number > 100kb (% of assembly)	931 (66%)
Mean scaffold size	76.9

Table 12: Approximate characteristics of genome assemblies of *D. baeri* genome, constructed from 10 pooled metacercariae

6.3.4 Mt genome content and annotation organization

The complete mitochondrial genome of *D. baeri* was a total of 14,480bp long, which contains 36 total genes being transcribed, 12 protein-coding genes, 22 tRNA genes and 2 rRNA genes (Table 1), consistent with findings in other species of digenea *D. baeri* lacks *atp8* gene(Olson *et al*., 2003). There is one NCR region within the *D. baeri* mt genome consistent with annotations of *D. pseudopathaceum,* however not consistent with *D. spathaceum,* which has two.

The arrangement of genes in *D. baeri* mt genome is similar to that of other published Diplostomidae genomes with NCR regions falling between rrnS (9764-10495) and cox2 (10525-11118) and the second falling between tRNAG (13598-13668) and cox3 (1-651) (Fig 1). In comparison to other species of Digenea the mt gene order is similar to Echinostomatidae, Heterophyidae and Opisthorchiidae, but distinct from species of African *Schistosoma* (T. H. Le *et al*., 2000). The nucleotide composition of *D. baeri* indicates a clear A-T bias, with 71.6 % of the entire genome regarded as A-T rich. The content of C is low (10.49%) and T is high (46.96%), the nucleotide composition is consistent with that of *D. spathaceum* and *D. pseudospathaceum (*Fig 26).

The tRNA genes of *D. baeri* ranged in size from 60 to 70 bp in length, indicating similar structure compared to of Diplostomidae (Brabec et al., 2015b) (supplementary figure 2). The large ribosomal RNA gene (*rrn*L) were located between trnaT (8636-8703) and *cox*2 (10525-11118) and separated by *trna*C (9696-9762), with the length of *rrn*L 973bp and *rrn*S 734bp. The non-recombining region (NCR) is located between trnaG and cox3 (Supplementary table 1).

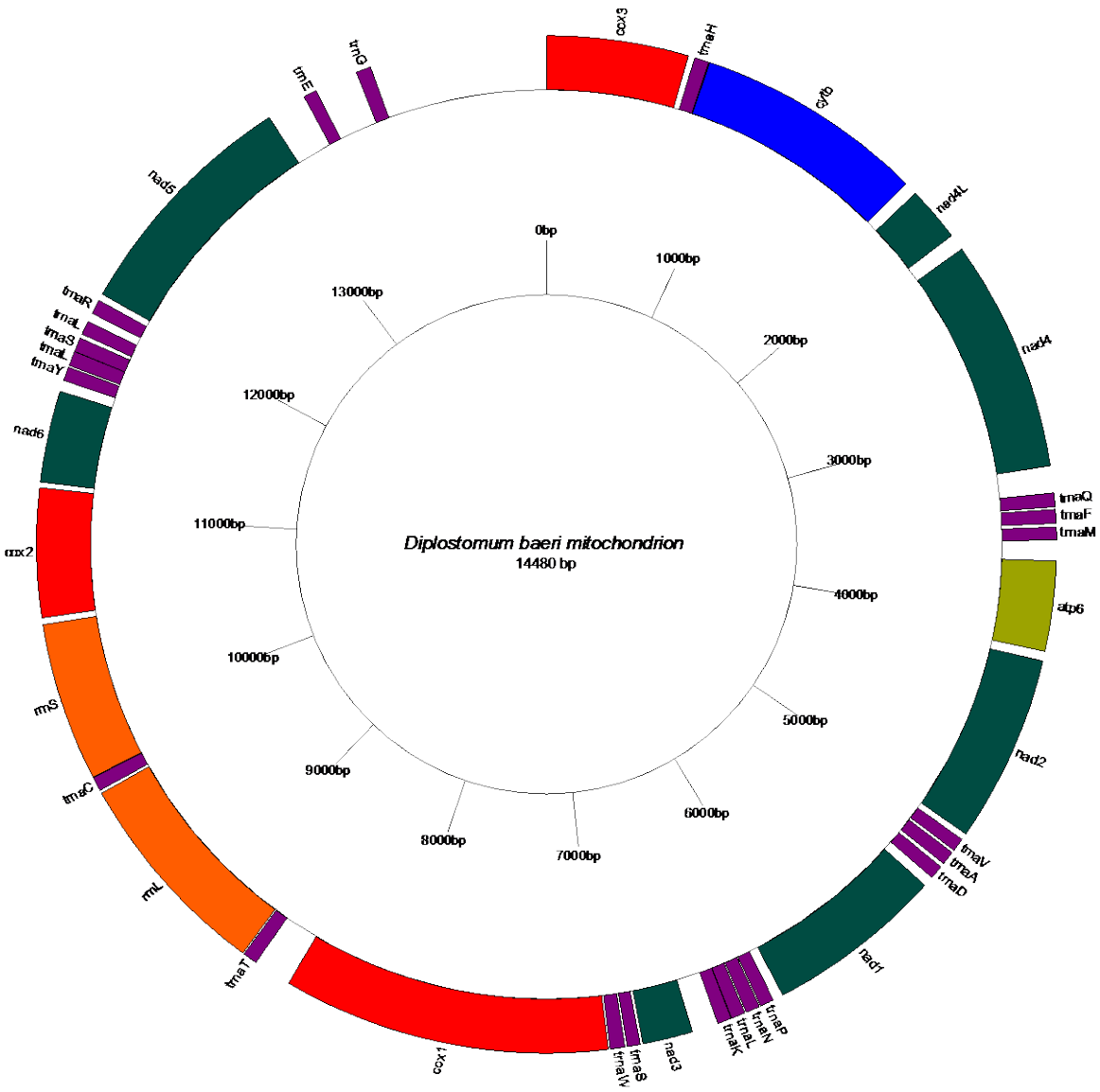


Figure 28; Circular visualization of mitochondrial genome with placement of labeled tRNA's, protein coding genes and ribosomal subunits.

A.

Species	A	T	G	C	G-C	Length (bp)
D. spathaceum	22.3	47.1	20.1	10.6	30.5	14,764
D. pseudospathaceum	22.4	47.9	19.8	9.9	29.5	14,099
D. baeri	23.7	46.9	18.9	10.5	29.4	14,480

B.

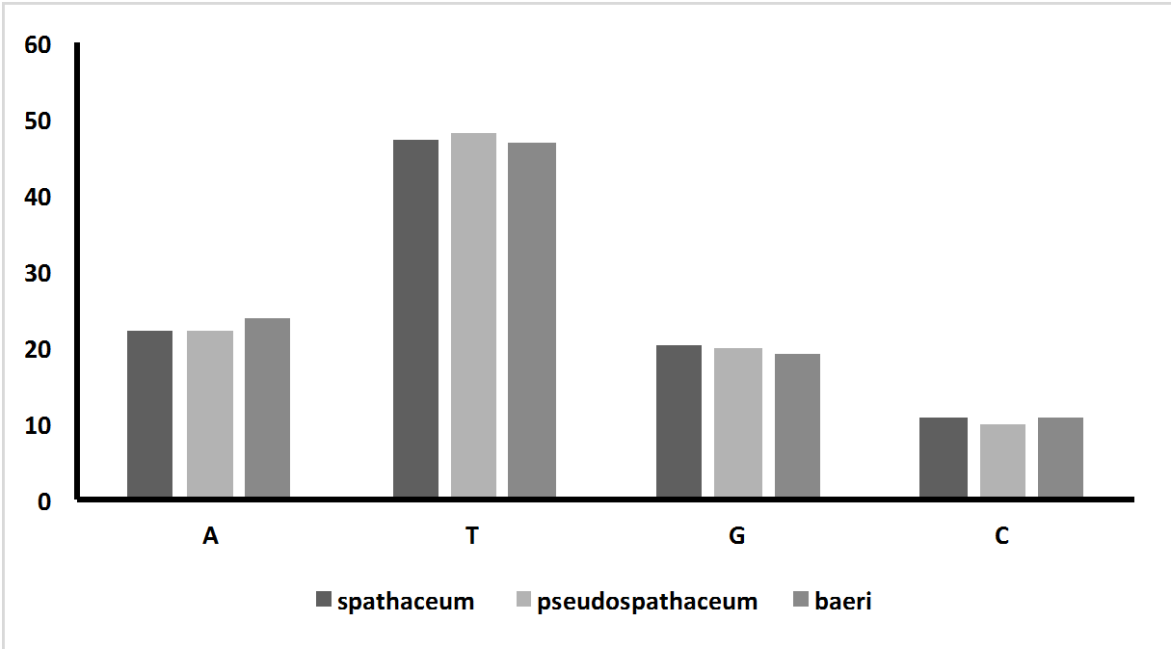


Figure 29; Comparison of mitochondrial genome between published *Diplostomum* species, A; table denoted percentage of A,T,C,G constitution of mitochondrial genomes across species. B; graphical chart in comparison of base percentage across *Diplostomum* species

6.3.5 Genetic divergence of *Diplostomum* Mt genome

The total difference between mitochondrial genomes across genus *Diplostomum* was 14.81% (2860 bp), which was comparable to the nucleotide difference between *D. baeri* and *D.pseudospathaceum* at 14.71% (1996 bp) and the nucleotide difference between *D. baeri* and *D.spathaceum* 16.13% (2189 bp) (supplementary table 2). Comparison of tRNA structure indicated variation in structure and RNA folding of Q,E,D,N,K,W,V,L2,E,G across *D. baeri, D.spathaceum* and *D.pseudospathaceum*.

At the protein coding level, differences between coding regions ranged from 7.89% to 19.09% between species *D. baeri* and *D.pseudospathaceum*. Comparisons between *D. baeri* and *D.spathaceum* showed slightly greater divergence with nucleotide differences across protein-coding genes falling within the range of 13.15% and 20.18% (Figure 27A). Analysis into specific gene conservation between *Diplostomum* species illustrated the most diverse genes were nad2 (82.07% conserved), nad4 (81.78% conserved), nad5 (81.29% conserved) and atp6 (82.36% conserved). The most conserved genes between across the genus were cox1 (88.42% conserved), cox2 (88.96% conserved) and cytb (86.98% conserved) (Fig 27B). These diversity characteristics were mirrored in other species of Digenea; *Echinostoma*, *Schistosoma* and Fasciolidae (Littlewood *et al*., 2006; Liu *et al*., 2016; Ma *et al*., 2016).

A

gene	S	Pi	K	position
cox3	106	0.1557	73.333	181
cob	199	0.13169	138.66667	751
nad4l	50	0.13856	35.3333	1863
nad4	285	0.19101	204	2192
atp6	106	0.17955	74.3333	3696
nad2-0	226	0.18171	157	4153
nad1	183	0.15485	126.66666	5312
nad3	58	0.17749	41	6571
cox1	245	0.1117	167.6667	6860
cox2	5	0.05464	3.333	10525
nad5	278	0.18179	198.3333	12044
nad6-1	61	0.20344	43.3333	13338

B

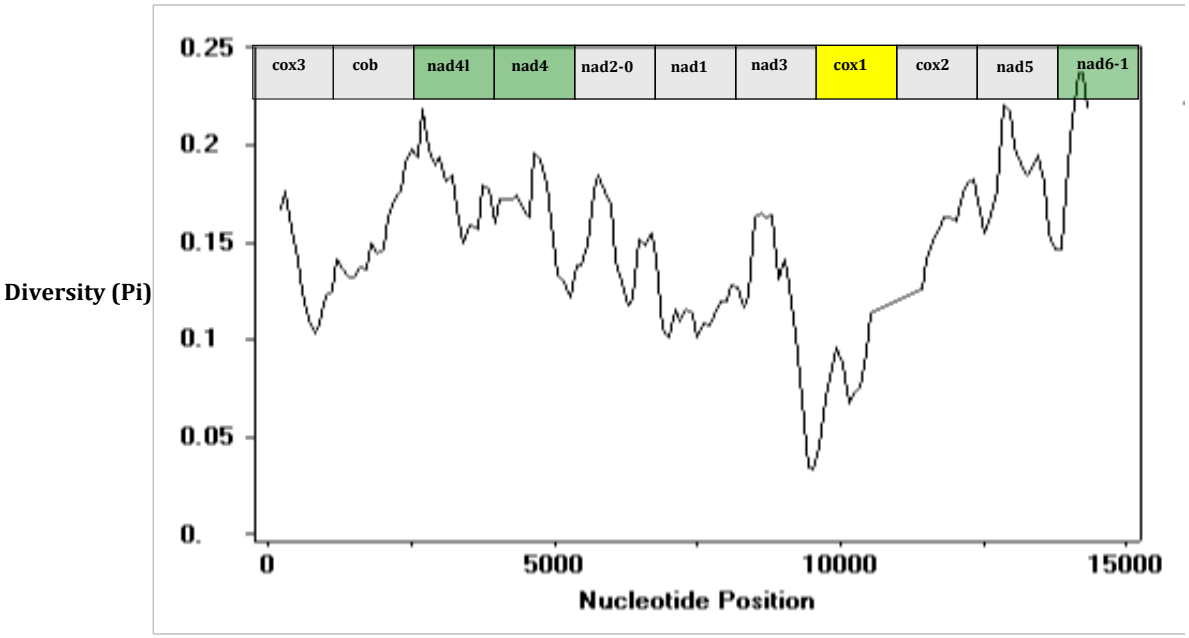


Figure 30; Divergence data across protein coding genes within *Diplostomum* mt genomes. A; Genetic divergence table S; segregating sites across *Diplostomum* species, Pi; observed diversity values, K; Average diversity existing per site. B; Graphical representation of divergence occurring across entire length of mitochondrial genomes between protein coding genes.

6.3.6 Phylogenetic analysis of mitochondrial genomes

The phylogenetic analysis across select Digenean mitochondria was performed on dataset of all 12 mt proteins. The mitochondrial genome amplified and annotated in this investigation grouped alongside two other *Diplostomum*- derived genomes, forming a Diplostomidae clade (Fig 28). The placement of this clade is flanked alongside the *Plagiorchiid*a species and *Schistosoma* clade, in accordance to previously investigated evolutionary relationships between digenea (Olson *et al*., 2003) (Figure 29). Nodal support was supported with strong posterior values in accordance with Bayesian inference

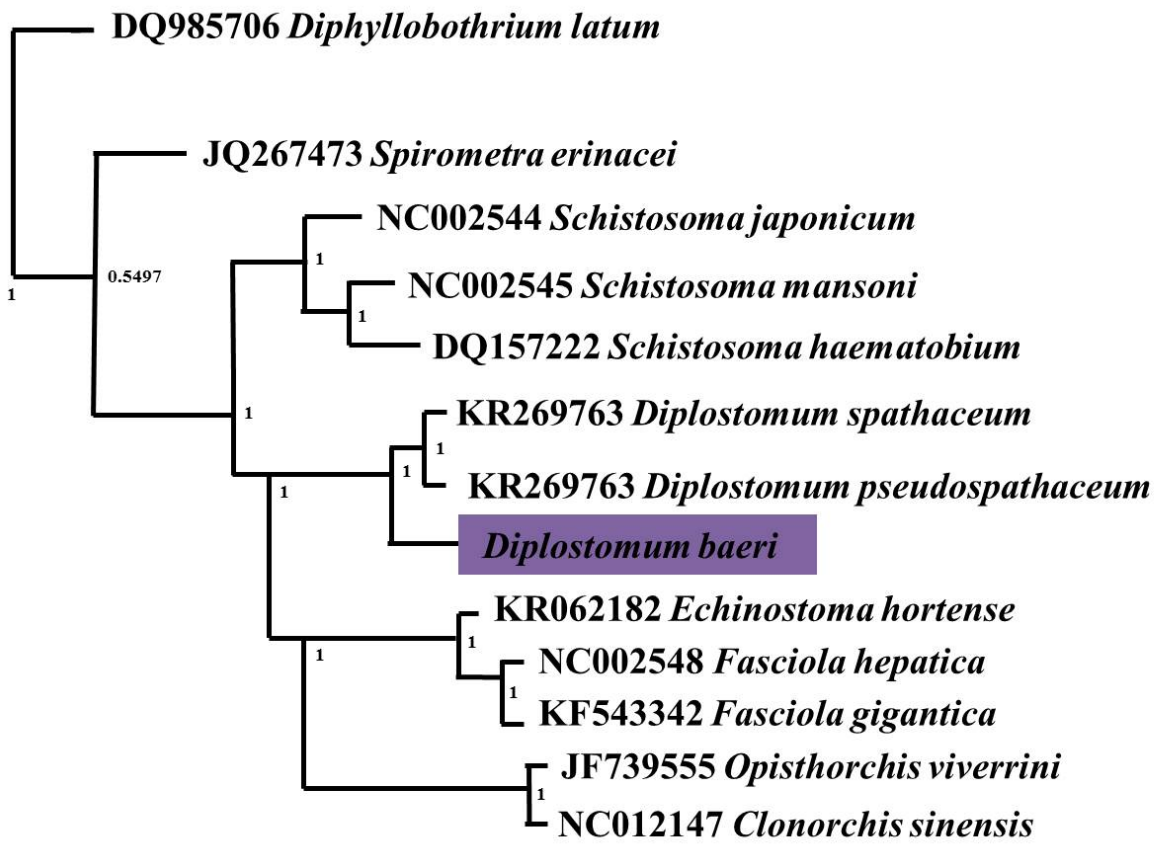


Figure 31; Phylogenic construction of protein coding genes in mitochondrial genomes across trematode species in accordance with Bayesian likelihood phylogenetic construction parameters. Highlighting of *D. baeri* within the phylogenetic construction is in purple

6.4 Discussion

6.4.1 Mitochondrial genome of *D. baeri*

This has been the first application of the NGS approach to completely characterize the mitochondrial genome of *Diplosotmum baeri* which has added to the limited genomic knowledge currently available regarding a widespread group of fish pathogens. The recent application of molecular techniques when studying *Diplostomum* has led to the discovery of an array of previously unrecognized species and species complexes existing within the genus (Blasco-Costa et al., 2014b; Georgieva et al., 2013c; Selbach et al., 2015a). The *cox1* barcode marker presented as problematic exhibiting low divergence within the *D. baeri (Blasco-Costa et al., 2014b), D. huronese* (Galazzo *et al.,* 2002*)*and *D. merg*i (Pérez-del-Olmo *et al*., 2014b) species. The molecular resources provided via this study can significantly improve the future aspects of molecular analysis using coalescent-based approach to species delimitation, particularly within the 'highly speciated' *D. baeri* species complex (Georgieva et al., 2013c; Selbach et al., 2015a). Finally directly corroborating with previous mitochondrial studies of *D.spathaceum* and *D.pseudospathaceum* (Brabec et al., 2015b), this study has also shown that the *cox1* gene is one of the most conserved protein-coding genes across the mitochondrial genome, having already been shown multiple species of parasitic flatworms (Le *et al*., 2002) and *Diplostomum* (Brabec et al., 2015b). Comparison of mt genomes of *Diplosotmum species* revealed that the least conserved markers are the dehydrogenase associated nad4 and nad5, with a high level of genetic divergence between the three closely-related species of *D. baeri, D.spathaceum* and *D.pseudospathaceum*. Use of the markers could enhance the efficacy of phylogenetic studies and enhance the ability to resolve some key problems in current *Diplostomum* taxonomic studies. The use of nad4 and nad5 markers may be advantageous for simultaneous species delimitation.. Within recent years the use of large scale environmental DNA (eDNA) (Bass *et al*., 2015) and coalescent NGS biodiversity studies (Preidis and Hotez, 2015) have allowed for the genetic identification of parasites that utilize freshwater stages within their life cycles. The genomic resource provided for *D. baeri* in this study provides short highly divergent DNA sequences which can be

utilized when amplifying bulk samples acquired from field-derived parasite samples and also free-living larval stages of the parasite through eDNA identification.

Mitochondrial data acquired from the study made it able to infer phylogenetic relationships of the order Diplostomida along with Plagoriorchiida which make up two of the largest clades within the order Digenea. The phylogenetic construction of mitochondrial genomes s infers the family *Diplostomidae* exists as a sister clade to the basal *Plagorichiddae*, this differs from the earlier phylogenies which saw a reversal of this, also current rDNA-based constructions (Olson *et al*., 2003). This can only be confirmed using limited mt genomes currently published and available, however if this is the case that Diplostomidae is an early diverging lineage from *Plagiorchiidae,* then the seemingly synapomorphic life history trait of cercarial penetration shared throughout *Diplostomidae* may be viewed as pleisiomorphie within the subclass Digenea. However addressing life history trait questions regarding *Diplostomum* would require additional genomic data, particularly within earlier divergent lineages, including; Spirorchidae, Brachylaimoidae and Clinostomidae). In terms of placement of *D. baeri*, it is placed alongside other published mt genomes of *D.spathaceum* and *D.pseudospathaceum*. Gene order comparisons of *D. baeri* within the neighboring *Plagiorchiida* clade see a minor alteration with a switched position of *trnP* and *trnN*. In comparison to neighboring *Schistomidae*, comprehensive gene order alterations are described, with variation in atp6 and nad2 along with multiple *trn* (V,M,W,S,N,I,K,A) positions, contrary to the minor alterations exhibited between *Diplostomidae* and *Plagiorchiida* reflected in phylogenetic reconstruction. The use of entire mt genomes in phylogenetic analysis can present important insights into the patterns and processes affecting mt genome evolution, however its use for species identification within large-scale population genetic studies still remains difficult in light of nuclear ribosomal and mitochondrial protein conflicts in phylogeny.

6.5 Conclusions

The present investigation to determine the complete mt genome of the important fish eye-fluke *D. baeri* reveals a close relationship between three species of *Diplostomum*. Comparisons of mt genomes across the genus has allowed for the identification of previously un-used mitochondrial markers that may give further answers to phylogenetic, biological and epidemiology questions, in *Diplostomum* and other freshwater trematodes.

7 Genomic mining for *D. baeri* antigenic vaccine targets

7.1 Introduction

7.1.1 Trematode parasite blood-stage infection

Freshwater trematode species frequently have complex life cycles utilizing multiple hosts within the freshwater environment before achieving definitive host infection. A key part of freshwater digenean life cycle is the successful invasion and infection of intermediate host (Criscione and Blouin, 2004). The parasite undergoes a significant transformation from free-living cercariae to attachment, penetration and blood vessel infection within an intermediate fish host (Curwen *et al*., 2012.). In the case of *Diplostomum* the parasite must reach the immunogenetically inert optical region, however to migrate to this region an extended period of time is spent within the host blood vessels (Brant *et al*., 2006). Migration of the parasite through blood vessels means it has to withstand a litany of host immunogenic responses. For the host to efficiently defend against parasite insult immune responses present a varied array of immunogenic counter measures. Immunogenic response is broadly divided into two systems, the innate immune system; an immediate response to infection utilizing "antigen presentation" a process by which the direct binding of MHC molecules to pathogen enacts a host immune response through multiple immune response cascades (Akira *et al*., 2006). The second form of immune reaction is the adaptive immune response involving pathogen-specific antibodies, which travel throughout the blood stream of the host recognizing pathogenic insult in a highly specific manner (Cooper and Alder, 2006). The capacity of freshwater parasites to modulate and evade both aspects of the host immune system underpins their infective success and longevity within a host population.

7.1.2 *Diplostomum* molecular host-parasite interaction

With its widespread infective ability within varied environments and intermediate fish species (Blasco-Costa *et al*., 2014; Grobbelaar *et al*., 2015; Kudlai *et al*., 2017; Moghaddam, 2015; Sultanov *et al*., 2014) *Diplostomum* could utilize some degree of immunomodulation to achieve such a high rate of infection (Betterton, 1974; Karvonen et al., 2003, 2006; Voutilainen et al., 2009). This tegumental surface layer is a host-interactive surface of the parasite, a syncytium that covers the entire worm bounded by a bilayer apical membrane, which acts as an interface between parasite and host immune response (Skelly and Alan Wilson, 2006). Primarily investigated within medically important *Schistosoma* species, this membrane has been highlighted as an important aspect of trematode survival whilst migrating through host blood vessels (Haas et al., 2007). The tegument has a large invaginated surface area populated by numerous secreted proteins performing a variety of essential functional interactions with the host such as nutrient uptake, environmental sensing and immune cell down-regulation (Skelly and Alan Wilson, 2006). One family of cell surface proteins implicated in the down-regulation of the host immune response is tetraspanins, a large superfamily of membrane-bound proteins present within all metazoans. Structurally they exist in a large extracellular loop that exists in a transmembrane state along the parasite's tegument (Hemler, 2001, 2003; Coceres et al., 2015). The membrane bound nature of tetraspanin superfamily proteins would include the involvement in disrupting any binding measures employed by host immune system such as MHC binding (Levy and Shoham, 2005b). This disruption of MHC binding would render the innate immune system impaired to parasite attack with MHC molecules being the initiating molecule of many innate immune cascades. Membrane-bound antigens can counteract direct binding immune cells, predominantly factors of the innate immune system, however to, counteract-blood bound specific antigens, freshwater digenean have a cocktail of secreted immune modulatory proteins to dampen parasite-specific antibodies. Included among these proteins is the Venom-Allergen-Like (VAL) family of proteins, these are members of the SCP (sperm coating protein) superfamily, which contains numerous commonly secreted

immunoregulatory proteins during, for example, *Schistosoma* invasion (Chalmers et al., 2008a) and gastrointestinal nematode infection (Hewitson et al., 2011). Venom allergen-like proteins are a large superfamily of proteins secreted by endo parasites during host invasion, the specific roles of which are poorly known. They are thought to play a role in immunoregulation. The proteins can be divided into two groups; group 1 are secretory proteins, released upon invasion into the host blood vessel, group 2 are structurally different with significantly less polarized sites within the amino sequences indicating the potential use as structural proteins within the membrane tegument (Chalmers and Hoffmann, 2012). With such a varied response by parasites to mitigate the host immune response there has, in recent years, been particular focus on the mediators released by parasites and the analysis of how these products may be responsible for successful parasite infection

7.1.3 Trematode antigen trans-membrane vaccine targets

Genomic study of membrane-bound antigenic products released by Digenea has now provided a more comprehensive catalogue of potential immunolomodulators which are involved in the critical interaction between parasite and host immune system (Hamilton *et al*., 2009). Identification of specific proteins involved in immune evasion mechanisms may help to develop vaccines to defend against Digenean infection. With their crucial role in host-parasite immune system interaction within the vascular system, tetraspanin proteins have been investigated as potential vaccine targets with polymorphic tetraspanin 2 (sm-TSP-2), being associated with reduced egg burden in *Schistosoma mansoni* infected mice (Loukas et al., 2007). By comparison, VAL proteins vaccine studies are still in their infancy (Chalmers and Hoffmann, 2012), primarily because of the large variety of proteins isolated but its direct IgE interaction has been confirmed in murine models (Abath and Werkhauser, 1996; Alba et al., 2014). Although antigens mentioned above have not been investigated within *Diplostomum* the action of immune evasion via blood stage parasite to increases survivability would be shared across freshwater Digenean parasites blood-vessel infective stage. With the high level of infective success demonstrated by various species of *Diplostomum* the characterization of antigens that

may increase efficacy of host invasion will be an important factor in understanding the infection dynamics of a widely prevalent parasite in freshwater systems. The genomic analysis of antigens may also allow for the discovery of potential vaccine targets within *Diplostomum* to be utilized within salmonid aquaculture.

7.1.4 Aims and objectives

The second part of the study was based on the genomic isolation of antigens within *Diplostomum.* This was performed utilizing *in silico* techniques to assess antigens as potential vaccine targets. The discovery and characterization of *D. baeri* membrane-bound antigens will also expand knowledge on structure and function of these peptides how they play a role in immune system interaction. The objectives in the second half of this study Includes:

- The genomic mining of TSP and VAL within *D. baeri* genome
- Translation of extracted antigen genes and structural analysis of TSP and VAL proteins
- Antigenic analysis of TSP and VAL proteins to find presence of direct binding sites of antigenic protein and host immune response complexes

7.2 Methods

7.2.1 Identifying antigens within Diplostomum baeri sequence

An exhaustive gene search was conducted to identify all tetraspanin (TSP) and venom allergen like (VAL) genes within draft genomes sequences of *Schistosoma mansoni* (assembly ASM23792v2). The position corresponding to the TSP and VAL locus was used to retrieve loci of the *Diplosotmum* genes. A two round BlastN search was performed using CLC genomics blast suite with cutoff *E* value of 10^{-15} was used against

the genome sequences. In the first round sequences of 10 TSP genes and 14 VAL genes, the extracted sequences were the only non-overlapping sequences given by the best hit (lowest *E* value). If the retrieved sequence was aligned with query sequence without any frame shifts or premature stop codons in the sequence and had proper RSS, the sequence was regarded as a potentially functional VAL or TSP gene. Once putative *Diplostomum baeri* antigens were obtained further scaffold BLASTn procedure was performed using the EXONERATE (https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate) command line program to manually verify and edit automated annotation procedures. The removal of indels from sequences was performed using GENEWISE (https://www.ebi.ac.uk/Tools/psa/genewise/) amino to nucleotide alignment program.

7.2.2 Functional analysis and modeling of trans-membrane antigen

Post annotation analysis was performed on translated sequences of *Diplostomum baeri* VAL and TSP antigen (Table 1). Structural analysis of exposed regions of antigen was done using BepiPred 2.0 (Jespersen et al., 2017) using epitope binding threshold of 0.5 and Exposed/Buried threshold according to NetsurfP threshold analysis. Potential Immunogloublin binding sites were discovered using IgPred server analysis (Tong *et al*., 2006), the protein was scanned using variable length parameters in 20 step-wise window analyses. To ascertain regions of the antigen that indicated increased potential for Major Histocompatibility Binding, translated sequences were analyzed using NetMHCII 2.3 software analysis (Jensen et al., 2018), using MHC II binding template from Murine H – 2 MHC-II molecules revealing MHC II-specific binding regions along exposed sections of trans-membrane antigen protein.

3D Amino homology models were created using SWISS-MODEL (Biasini et al., 2014) protein structure server, with optimization performed using MODELLER 9.1 (Eswar et al., 2006) and visualized and produced using MacPyMOL (Delano, 2007) software.

7.3 Results

7.3.1 Gene identification and annotation

The total number of tegument-associated antigens found within the *D. baeri* genome mapping was 7, these included 3 tetraspanin-like antigens (TSP) and 4 Venom allergen-like antigens (VAL) (Table 12). The annotation of TSP antigens within the *Diplostomum* genome were checked using BlastN analysis in NCBI Blast suite with *Diplostomum* genes showing similarity to previously annotation TSP antigens within *Schistosoma; D. baeri*TSP-2; 85% ident score *S. mansoni* TSP-2 (Accession AAN17276) E value: 7e-138, *D. baeri*TSP-23; 76% ident score *S. mansoni* TSP-23 (Accession AGA82197) E value: 1e-67 and *D. baeri*TSP-31; 80% ident score *S.haematobium*TSP-31 (Accession XP012795895) E value: 2e-53. The *D. baeri* genes that show similarity to previously annotated VAL antigens within *Schistosoma*; *D. baeri* VAL-5; 85% ident score *S.mansoni*VAL-5 (Accession XP01864402) E value 2e-145, *D. baeri*VAL-11; 87% ident score *S.mansoni*VAL-11 (Accession XP018650104) E value 0.0, *D. baeri*VAL-27; 79% ident score *S. haematobium* VAL-27 (XP012798637) E value 6e-102 and *D. baeri*VAL-18; 71% *S. mansoni* VAL-18 (Accession XP018652917) E value 2e-76. B

Antigen	Epiptope	Amino Seq	Binding Site
Diplostomum TSP-2	MHC II	**PKDYGENPPTSCSKDG**	157-173
	Immunoglobulin	**GENPPTSCSKDGVQFTQGC**	145-165
Diplostomum TSP-23	MHC II	**MTGALDKPTRVITEFMDL**	106-132
	Immunoglobulin	**AREAIIVVGVIILIVSFLGCCG**	21-41
Diplostomum TSP-31	MHC II	**AGCPPCLALCCCDTPDCE**	34--52
	Immunoglobulin	**KDDLLPSDPAGCPPCKLAK**	7--27
Diplostomum VAL-5	MHC II	**GFQRWLNEYKNWDFFNRLC**	118-137
	Immunoglobulin	**YVKQRCNHTNERQLNRTPA**	211-230
Diplostomum VAL-11	MHC II	**ISPDGHLVDNSKKLQESIHS**	226-246
	Immunoglobulin	**RGFSDNTDDKGXNKKLYVGQ**	172-192
Diplostomum VAL-27	MHC II	**WAGAKDIQSWLAEYLKNYDFYTRTCRG**	97-124
	Immunoglobulin	**SEVRNGQLFGQPRAVSIK**	41-59
Diplosotmum VAL-18	MHC II	**QPYESRPHDRGDDGDGAVY**	161-179
	Immunoglobulin	**AEVDKQPYESRPHDRGDDGDGAV**	153-173

Table 11; Description of putative antigen genes annotated and extracted from *D. baeri* genome. Epitope; defines the immunological binding potential of "MHC II" domains and "Immunoglobulin" domains. Amino Seq: denotes raw sequence of protein that exhibits increased potential to bind with immunological factor. Binding site; represents positioning of binding domain within protein.

7.3.2 D. baeri Venom allergen-like protein analysis

7.3.2.1 VAL protein structure

The VAL protein sequences annotated from *D. baeri* genome varied in structure with, Venom allergen-like antigens are broadly divided into 2 groups, one of the VAL proteins extracted from *D. baeri* present variable cysteine residue within the C-terminal regions.

Group 1 VAL-18 protein indicated trematode specific ‘M’ group ligand existing within terminal region. In all VAL proteins extracted for *D. baeri* ‘trematode’ specific cysteine group was present before the first helix and within the C-terminal (Cys26-Cys195) (Figure 30), this group is confirmed as trematode-specific across multiple species of parasites (Chalmers et al., 2008b). Consensus group 1 VAL protein construction revealed increased cysteine presence within group 1 3D construction (Figure 7A) with 8 cysteine groups within the amino sequence compared to only 3 in group 2 *D. baeri* VAL proteins.

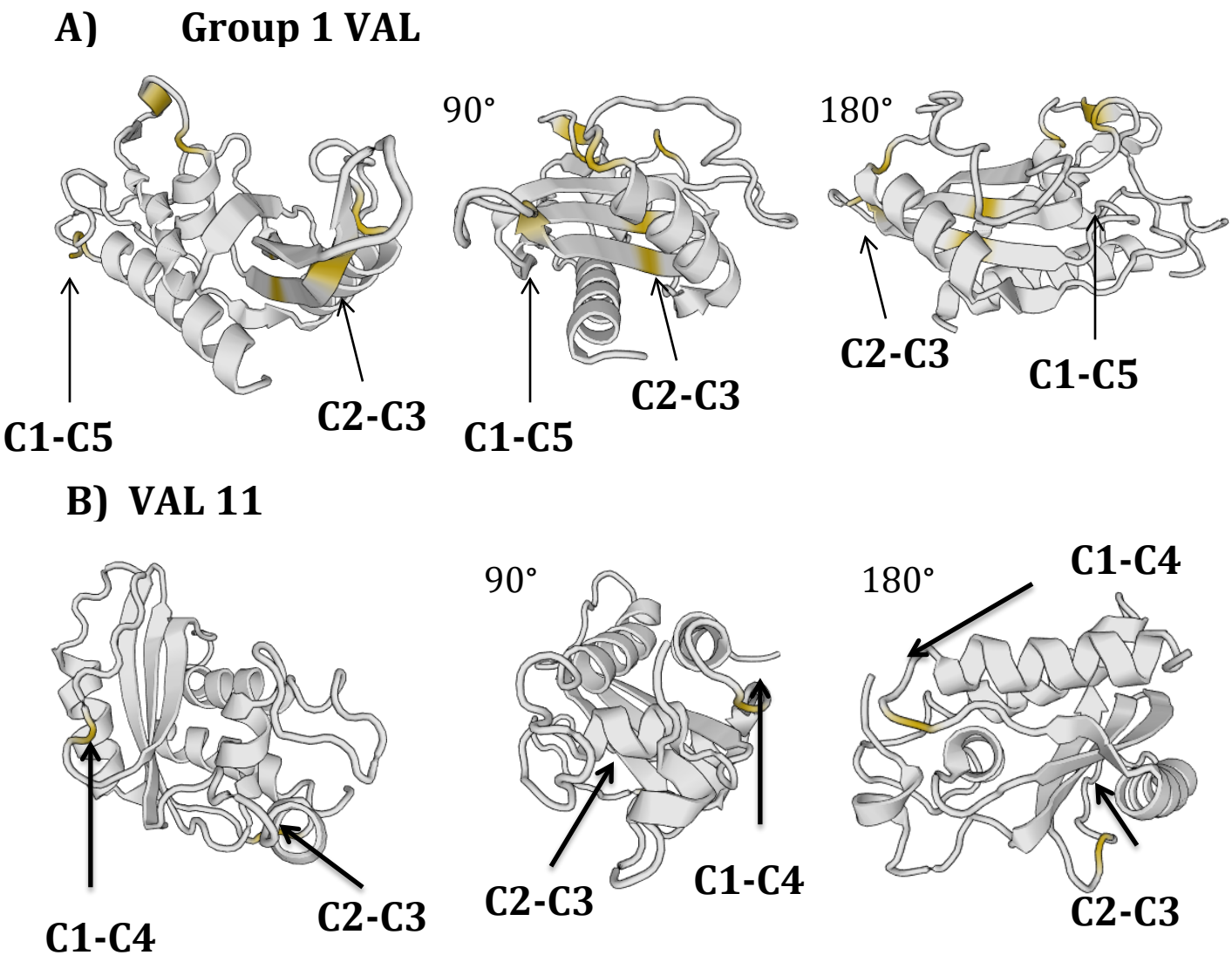


Figure 32; Diversity of structural domains between group 1 VAL *Diplostomum* (27,5,11)consensus protein structure with exposed cysteine motifs within structure highlighted in yellow. (A) Homology model of Group 1 *Diplostomum* VAL proteins, group 1 C-terminal region is highlighted in yellow exposing potential disulphide bonds across protein structure. C-terminal regions are arrowed and annotated across structure (B) Homologous structure of VAL-11 in *D. baeri*, reduced C-terminal regions are highlighted in yellow and labelled across structure

7.3.2.2 VAL protein phylogenetic construction

A total of 74 trematode VAL domain amino acid sequences were aligned, with phylogenetic construction revealing divisions between trematode members of group 1 and group 2 VAL proteins (Figure 31). *D. baeri* VAL proteins placed alongside respective members of similar VAL proteins with *S.mansoni* VAL 18, 11, 27 and 5.. With Group 1 proteins displaying a lack in resolution within clades containing VAL-5, 15, 26 and 28, *D. baeri* VAL-5 protein is placed alongside a mixed VAL-5 and VAL-15 clade. Group 2 VAL proteins placed alongside other group 1 proteins throughout phylogenetic tree construction including *D. baeri* VAL-11 deriving from *D. baeri* genome. Bootstrapping supports remained consistent throughout the phylogenetic construction maintaining well-supported clade division.

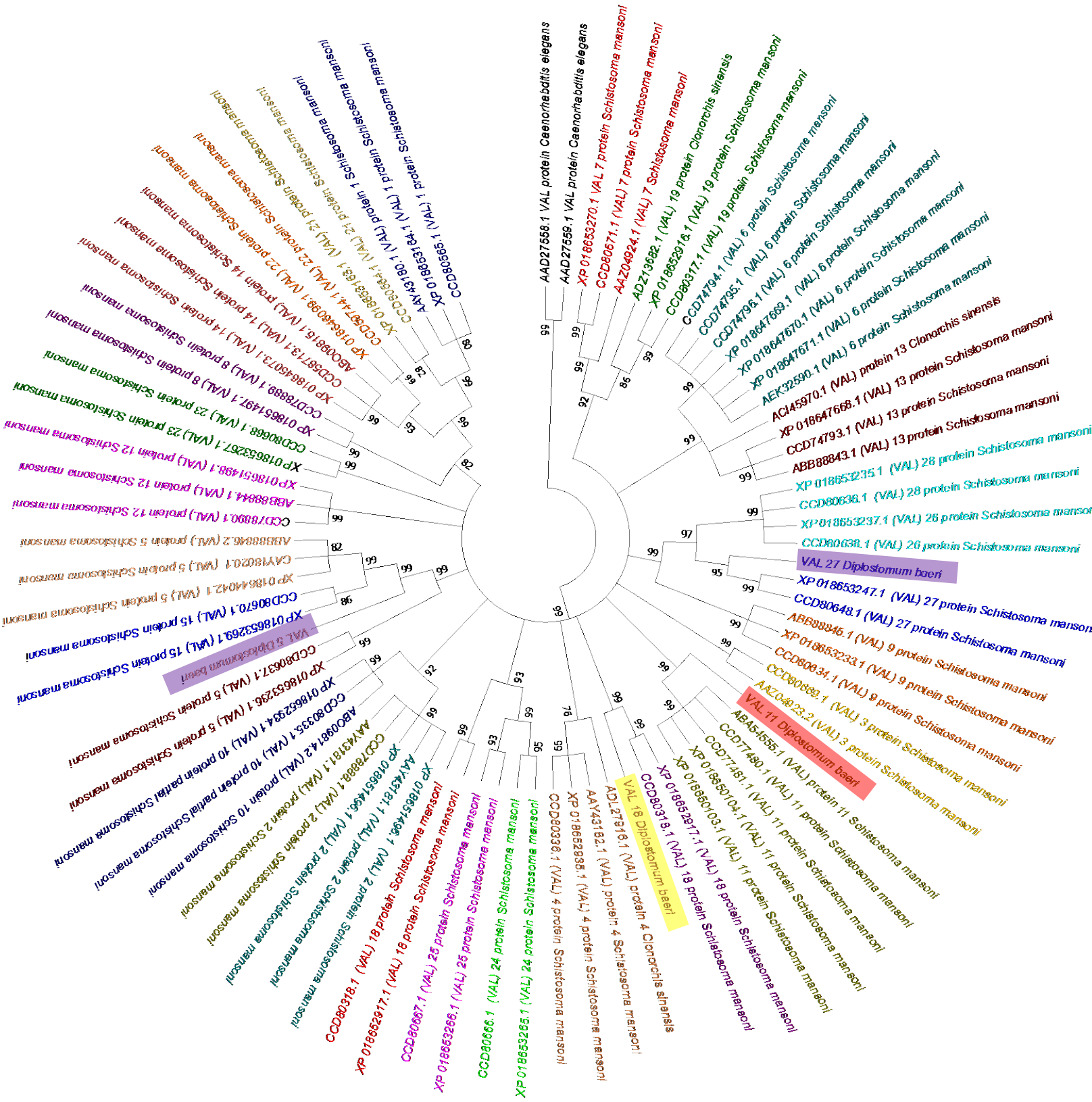
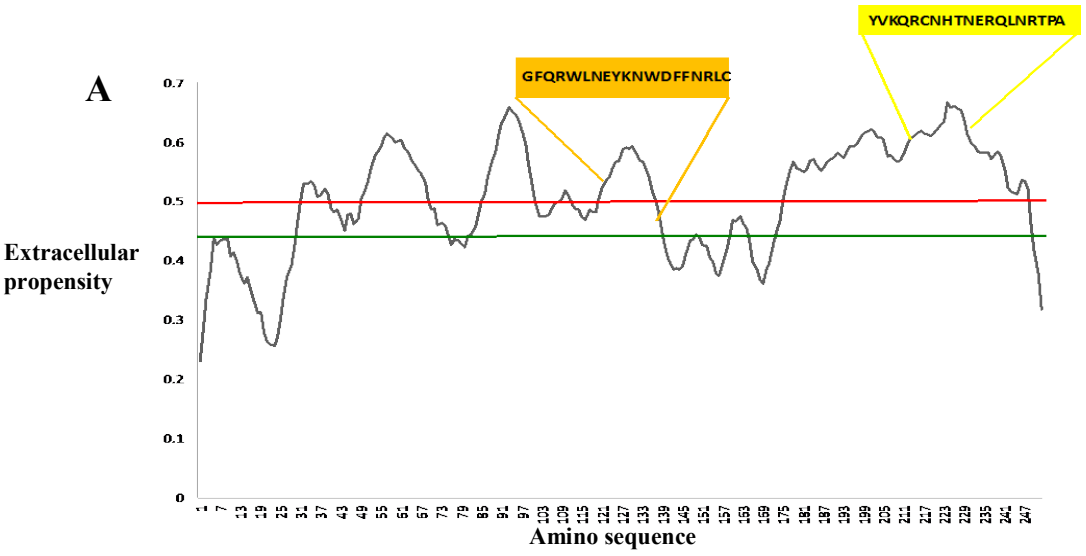


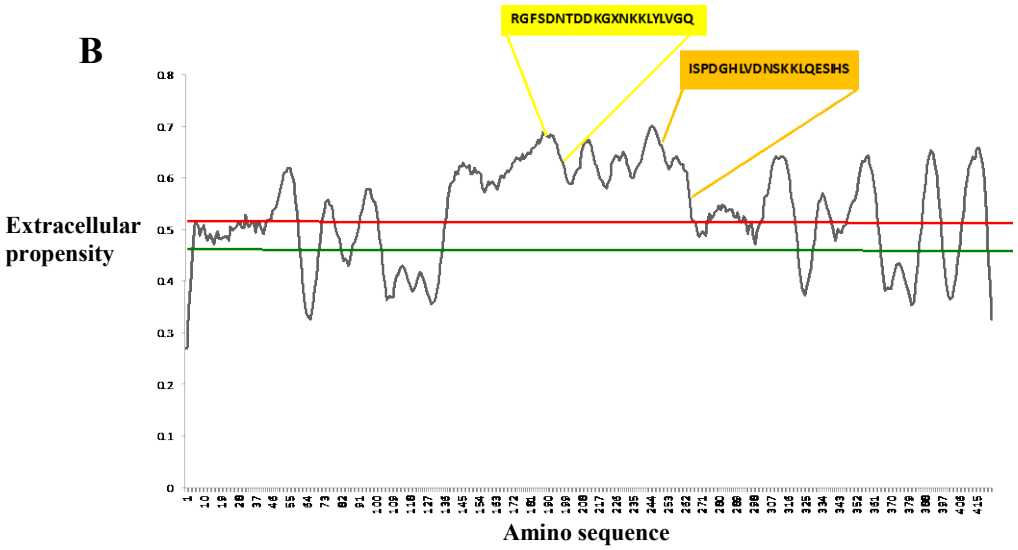
Figure 33: Phylogenetic construction of venom allergen-like (VAL) proteins across trematode species *Schistosoma mansoni*, *Clonorchis sinensis* and *Diplostomum baeri*. The VAL proteins specific to *Diplostomum baeri* are highlighted on the tree as VAL-27, VAL 11, VAL 18 and VAL 5

7.3.2.3 Antigenic propensity of VAL proteins

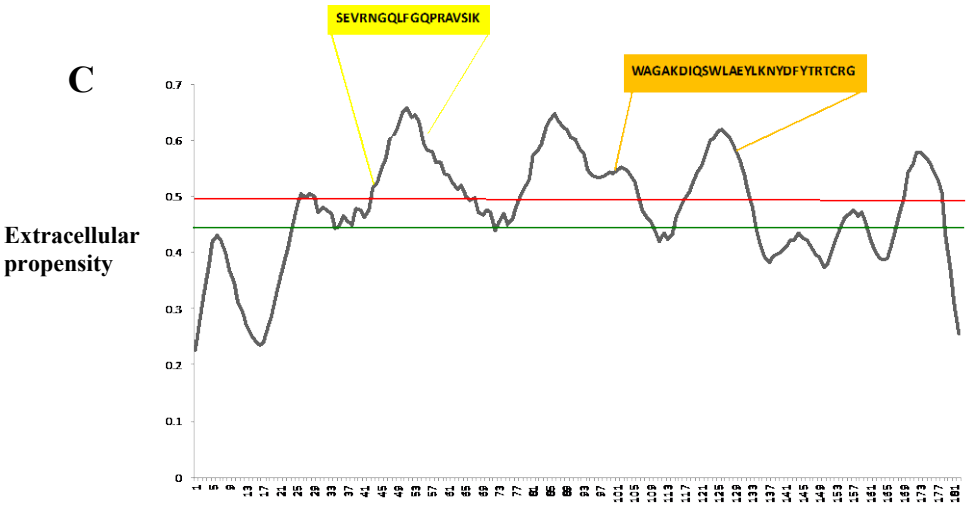
All VAL proteins annotated from *D. baeri* exhibited regions of the protein indicating antigenic propensity. Antigenicity graphs have been combined with epitope analysis to show potential binding epitopes of immunoglobulin binding domains and MHC class II potential binding domains. All VAL proteins showed regions that could indicate potential epitope binding regions. *D. baeri* VAL-5 predicted MHC binding domain within extracellular region 118-137 (GFQRWLNEYKNWDFFNRLC), immunoglobulin epitope binding domain was located between 211-230 amino sequence (YVKQRCNHTNERQLNRTPA) (Figure 32A). *D. baeri* VAL-11 predicted MHC binding domain within extracellular region 226-246 (ISPDGHLVDNSKKLQESIHS) and immunoglobulin predicted binding domain between 172-192 amino domain (RGFSDNTDDKGKNKKLYVGQ) (Figure 32B). *D. baeri* VAL-27 predicted MHC binding between 97-124 amino domain (WAGAKDIQSWLAEYLKNYDFYTRTCRG) and immunoglobulin binding domain predicted between 41-59 amino domain (SEVRNGQLFGQPRAVSIK) (Figure 32C). *D. baeri* VAL-18 was the only VAL protein to indicate similar epitope binding domains of both MHC and immunoglobulin. The predicted VAL-18 MHC binding location between 161-179 amino domain (QPYESRPHDRGDDGDGAVY) and immunoglobulin-binding domain 153-173 (AEVDKQPYESRPHDRGDDGDGAV) (Figure 32D).



VAL-5



VAL-11



VAL-27

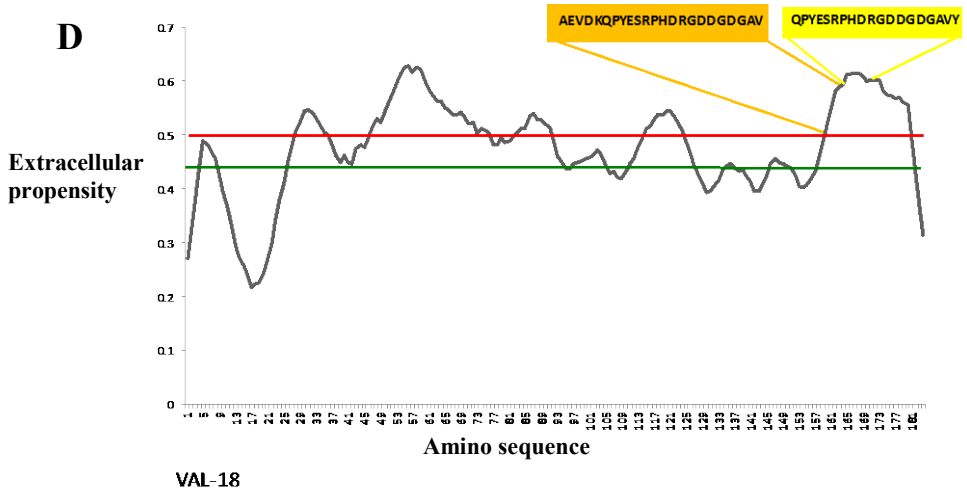


Figure 34; Epitope binding plot denoting structural extracellular propensity, Red line indicates out membrane, Green line represents intracellular membrane. Immunogenic epitope binding domain is highlighted with Orange for Immunoglobulin binding domain and yellow for MHC II binding domain. Graphs are labelled A; VAL-5, B; VAL-11, C; VAL-27, D; VAL-18.

7.3.3 D. baeri Tetraspanin protein analysis

7.3.3.1 Tetraspanin structural analysis

Tetraspanin superfamily proteins annotated from *D. baeri* genome consisted of putative TSP-2, TSP-23 and TSP-31 (Table 1). Structural analysis of TSP proteins indicated the consistent presence of extracellular loop structure across all three TSP types in *D. baeri*; this is a shared structural motif amongst TSP proteins. Another conserved aspect of TSP proteins is the presence of a cysteine-cysteine-guanine (CCG) region within the extracellular loop (Figure 33), the CCG region is conserved across all studied TSP proteins and is present within *D. baeri* putative TSP structure. Charge analysis across the 3D structure of the TSP proteins is highlighted within colored regions of the 3D structure of all *D. baeri* TSP (Figure 34), this analysis consistently indicated the presence of charged amino regions being located on the extracellular loop domains, with the extracellular loop being the primary binding point between parasite tegument and the host immune system an increased charged amino potential within this domain would enhance binding likelihood.

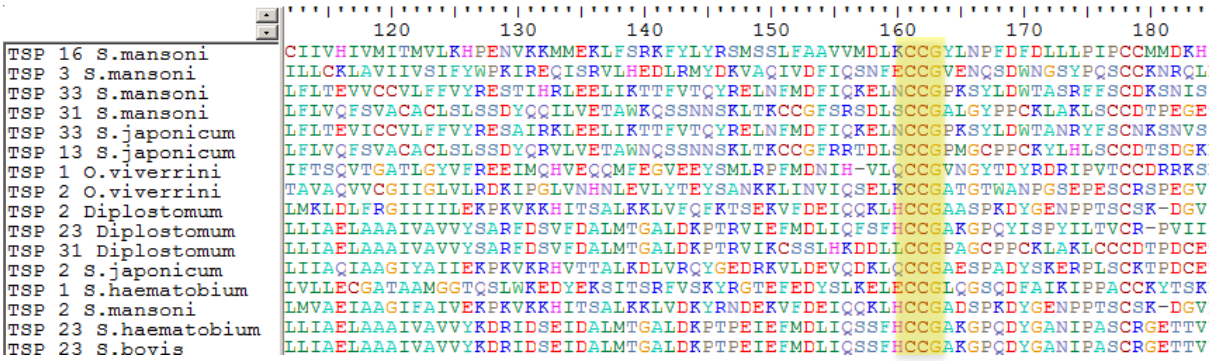


Figure 35; Multiple translated sequence alignment of TSP proteins from *S.mansoni, S.japonicum, S.haematobium, S.bovis* and *O.viverrini*. Highlighted yellow region denotes 'CCG' conserved region across TSP proteins.

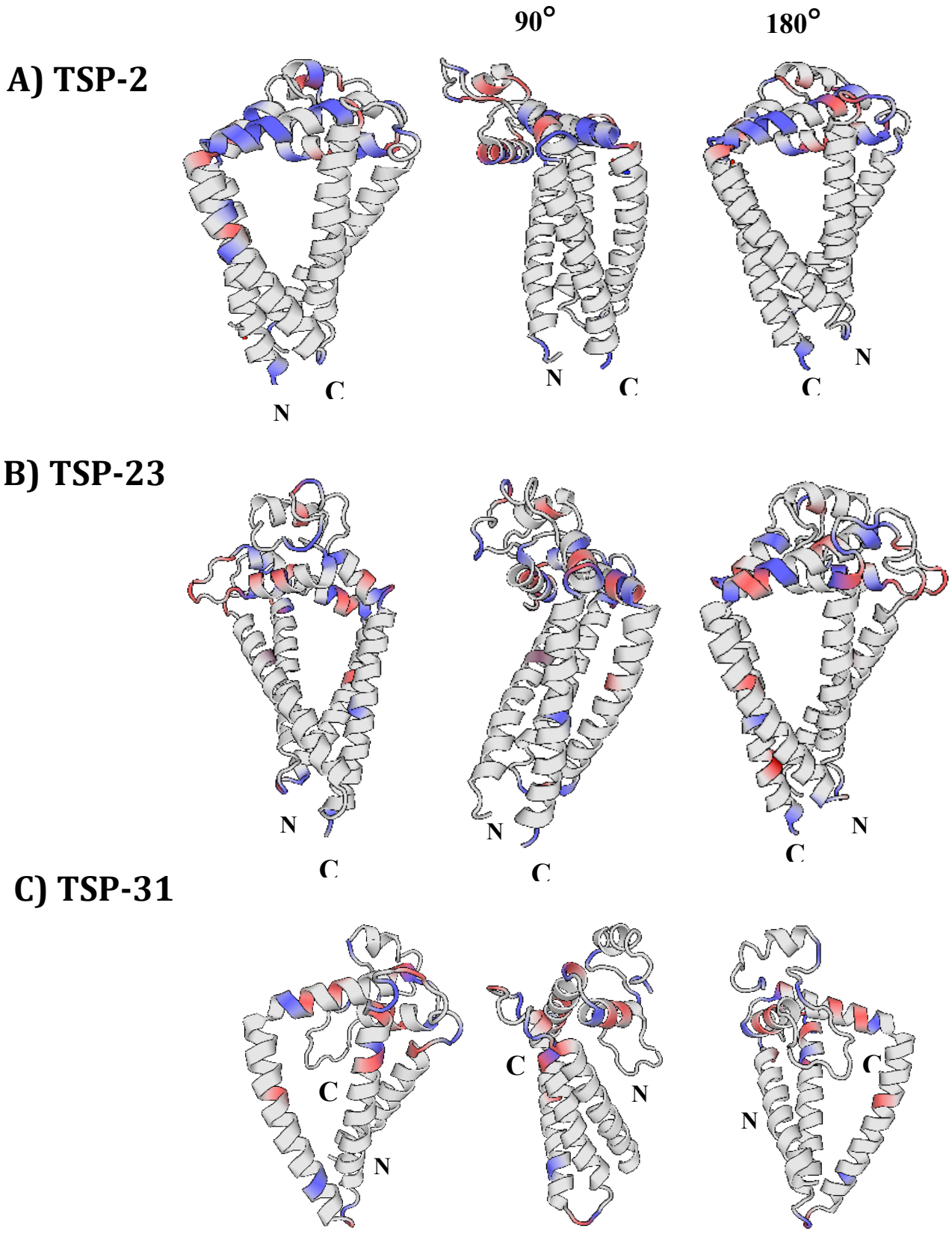


Figure 36; Graphical representation of 3D matrix reconstruction of the extracellular loop from protein alignment of TSP 2,23 and 31 from *D. baeri* genome. All protein structures are rotated 90 and 180 degrees with N and C terminal regions of the proteins labelled on diagram. Highlighted colours represent polarity of individual amino, using red for negative charge and blue for positive charge.

7.3.3.2 Phylogenetic construction tegumental proteins

Phylogenetic construction of tetraspanins (TSP) proteins (Figure 35) was supported via strong posterior values with strong nodal support. *D. baeri* derived TSP proteins placed alongside respective *S.mansoni* (TSP-2 and TSP-23) and *S. haematobium* (TSP-31) previous putative matches using nBLAST workflows. All *D. baeri* TSP proteins presented as sister group alongside *Schistosoma* main clade not indicating high enough similarity to place within the main clade of TSP-2, TSP-23 and TSP-31.

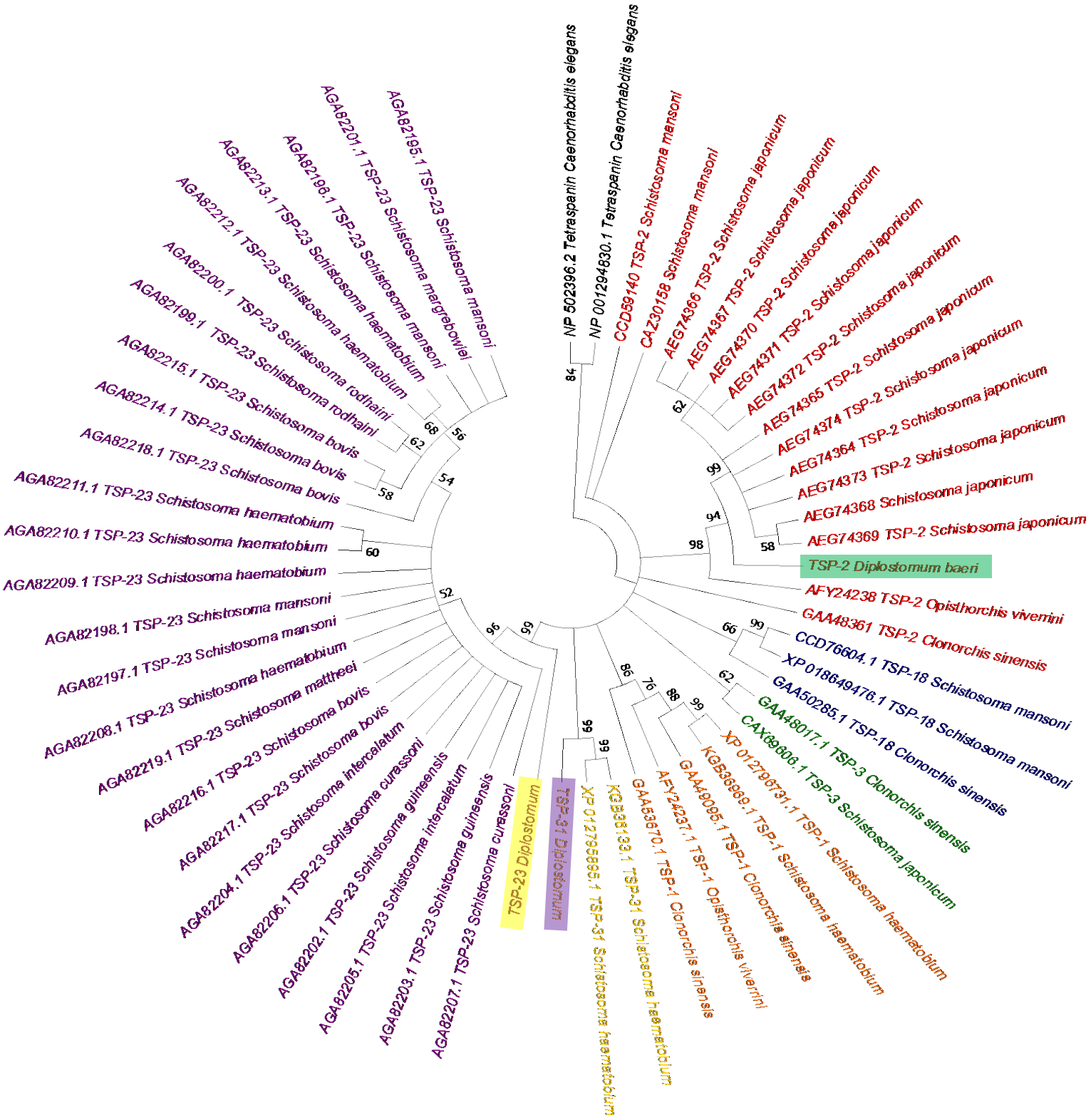


Figure 37; Phylogenetic construction of tetraspanins (TSP) proteins across trematode species *Schistosoma mansoni*, *Clonorchis sinensis* and *Diplostomum baeri*. The tetraspanins specific to *Diplostomum baeri* are highlighted on the tree as TSP-2, TSP-31 and TSP-31.

7.3.3.3 Antigenic propensity of TSP proteins

All TSP proteins annotated from *D. baeri* annotation exhibited regions of the protein indicating antigenic propensity. Antigenicity graphs have been combined with epitope analysis to show potential binding epitopes of immunoglobulin binding domains and MHC class II potential binding domains. *D. baeri* TSP-2 had a total length of 240 aminos with the epitope binding region for MHC II being located 157-173 domain (PKDYGENPPTSCSKDG) and the region exhibiting increased immunoglobulin binding located between 145-165 (GENPPTSCSKDGVQFTQGC) (Figure 36A). *D. baeri* TSP-23 was a total of 223 aminos in length with the MHC II epitope binding region located at 106-132 (MTGALDKPTRVITEFMDL) and the immunoglobulin-binding domain being located between 21-41 (AREAIIVVGVIILIVSFLGCCG) (Figure 36B). *D. baeri* TSP-31 has a total length of 110 aminos in length with the region exhibiting MHC II binding propensity located between 34-52 (AGCPPCLALCCCFTPDCE) and the immunoglobulin-binding domain being located between 7-27 (KDDLLPSDPAGCPPCKLAK) (Figure 36C).

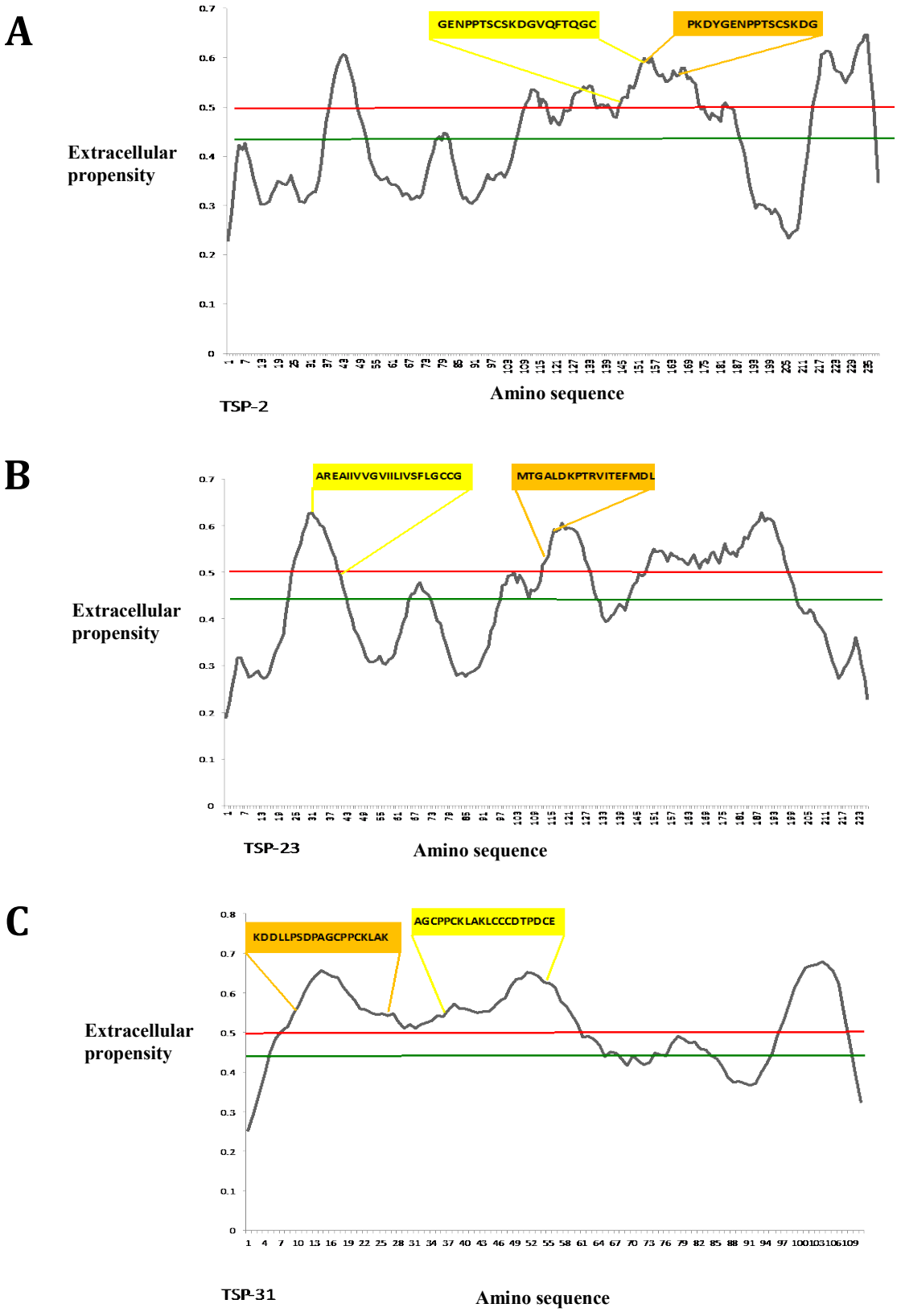


Figure 38; Epitope binding plot denoting structural extracellular propensity, Red line indicates out membrane, Green line represents intracellular membrane. Immunogenic epitope binding domain is highlighted with Orange for Immunoglobulin binding domain and Yellow for MHC II binding domain. Graphs are labeled: A; TSP-2, B; TSP-23, C: TSP-31

7.3.3.4 Extracellular modeling of TSP proteins

Modelling of *D. baeri* tetraspanins was performed using Moddeller protein-modelling software to predict intra and extracellular regions of tegumental proteins. All *D. baeri* TSP proteins indicated both extracellular and intracellular regions, with TSP-2 and TSP-23 having the one small extracellular loop followed by a larger second extracellular loop (Figure 37). TSP-31 showed only a partial extracellular loop with only one end embedded within the membrane, this may mean only a partial protein sequence was extracted from *D. baeri* genome or TSP-31 may only be utilized differently within the TSP tegumental web than TSP-2 (Figure 37A) and TSP-23 (Figure 37B). Both TSP-2 and TSP-31 (Figure 37C) had an N-glyco motif within the TSP structure. Epitope binding domains examined in previous epitope binding graphs showed a MHC II and Immunoglobulin potential binding domains are located within extracellular regions of TSP-2, TSP-23 and TSP-31.

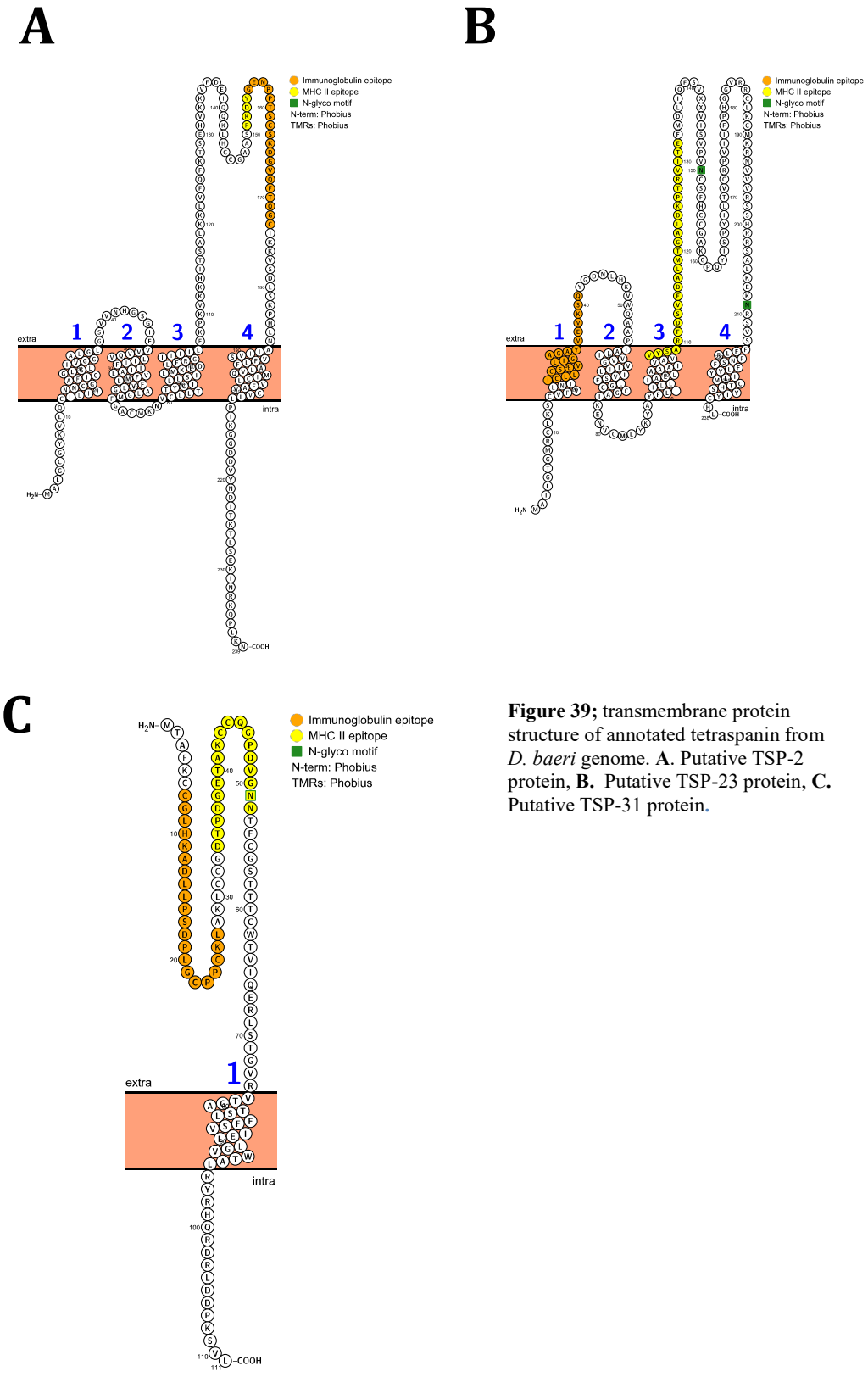


Figure 39; transmembrane protein structure of annotated tetraspanin from *D. baeri* genome. **A**. Putative TSP-2 protein, **B.** Putative TSP-23 protein, **C.** Putative TSP-31 protein.

The use of next generation sequencing technology has allowed for a rapid increase in the knowledge of parasitic genomic processes and to help understand biology of certain parasitic organisms. DNA microarrays, proteonomic technologies and whole-genome construction have allowed the discovery and downstream comparisons of numerous proteins and genes associated with pathogenic capabilities of parasites. Most of the studies have aimed to identify membrane-bound target proteins within the parasite tegument with numerous studies implicating them as potential vaccine targets (Loukas et al., 2007; McManus and Loukas, 2008; Pérez-Sánchez et al., 2008). The advent of high volume throughput and computational analysis has aided the annotation of genes of new parasites. Here the manual annotation of antigens within one of the most widespread freshwater parasites in the world has provided potential markers for future studies in host-parasite infection and freshwater molecular epidemiology. This study has been the first to identify and annotate sequences for tegument-bound proteins within the *D. baeri* genome. Tetraspanin-like proteins within *Diplostomum* show conservation in structure compared to *Schistosoma* counterparts, with extracellular loop-like proteins exhibiting similar "CCG" structural domains, a trademark in the amino profile of tetraspanin proteins. This may exhibit a crucial component in the structure of tetraspanin proteins with the motif located on the extracellular loop, the most likely point of host interaction. 3D structural protein construction displayed increased polarity of individual extracellular amino acids that would indicate the increased antigenic propensity of these regions to interact with extracellular components of the host immune system. The discovery of multiple tetraspanin-like proteins within the *Diplostomum* indicates that the proteins facilitate infection in similar ways to those associated with other digenea. During parasite invasion into intermediate host tetraspanins interact with one another to form a 'web' where the partnering of tetraspanin proteins can provide a more efficient interaction surface with extracellular proteins (Levy and Shoham, 2005a). The interactions of tetraspanins within the infective stage of *Diplostomum* is not limited to self-self interactions, epitope binding prediction plots indicate specific sections of the proteins where host proteins can interact with TSP genes. One of the crucial host-parasite molecular relationships discovered on the *Diplostomum* TSP proteins include MHC II

potential binding domain on the antigen. As part of the analysis into the modeling of the antigenic protein (Figure 35), indicates potential binding epitope sites as being within the tegumental membrane. It is not the case that immunoglobulin or MHC binds to these points on the antigen however by nature the tegument is shedding and reformed as the parasite blood stage passes through the blood stream of its host (Simpson et al., 1981; Kruger and Joubert, 1990; Van Hellemond et al., 2006). The shedding of the parasite tegument can expose intracellular antigens to further immune cell binding and interaction, making the parasite increasingly susceptible to destruction if recognized by the immune system. The tegument is not a smooth surface; the bilayer exists with vesicles and invaginations (Simpson et al., 1981; Gobert et al., 2003; Pérez-Sánchez et al., 2008) to increase the surface area and therefore increasing the antigen repertoire. This means that intracellular antigen modeling can only be consider hypothetical-computational modeling of genomic mined tetraspanins, with binding epitopes predicted through computational analysis.

The MHC is a vital part of the recognition and initiation of the innate immune response system that can recognize extracellular domains of pathogens to enact T cell host response (Simpson, 1988; Borghans et al., 2004). The Immunoglobulin binding domain is present on all TSP genes analyzed with epitope prediction showing clear binding potential within the extracellular domains of the TSP protein. Epitope prediction highlights the direct interaction that TSP proteins have with proteins associated with host response indicating the potentially vital role in successful infection.

The newly identified *D. baeri* specific VAL proteins indicated conserved structural characteristics in comparison with other members of trematode class. The 4 VAL proteins annotated within *D. baeri* were named due to similarity in matching with VAL proteins in other species. One VAL protein is classified as sharing pairwise similarity with group 1 VAL protein (VAL 18 *Diplostomum*), group 1 modeling has indicated that it is expressed at high levels during host invasion and that it may play a key role in the disruption of the host innate immune defence mechanisms such as the complement pathway. With intriguing symmetry a different set of ‘group 1’ proteins have been implicated in miracidia/sporocyst parasitism correlating with VAL 27 discovered in *D.*

baeri and VAL 5 associated with adult egg secretions with definitive host (Kelleher et al., 2014). Extraction of VAL proteins from the *D. baeri* genome revealed substantially less genes than *Schistosoma* counterparts. Venom allergen-like proteins are unlike Tetraspanins in the sense some of the family are actively secreted and do not reside purely within the tegument of the parasite, however the protein still exhibits potential binding domains with immunoglobulin. The epitope-binding domains associated with immunoglobulin within this study represent potential domains that show increased propensity for immunoglobulin binding.

Human hosts of Digeneans display a great diversity of immunoglobulin with 9 classes however salmonids have only three immunoglobulin classes (IgM, IgG and IgT)

Salmonid Ig compounds are involved in differential aspects of host immunity; IgD and IgT have been associated with mucosal tissue immunity, with IgT demonstrating up-regulation in response to pathogen infections and could be implicated as protective against *D. baeri* infection. There is strong proteomic evidence to suggest that group 1 VAL's are secreted by several trematode species during parasite invasion, with the presence of group 1 protein within *D. baeri* genome (VAL 18) the potential use of VAL 18 could be implicated as a major aspect of *D. baeri* invasion.

8. General Discussion

8.1 Introduction

Research into the parasitic infection of salmonids has gained ground as an area of research in recent years due to increasing global reliance of fish farming. Although farmed salmonid production is seeing an increase worldwide the conservation status of wild populations of salmonid remains varied. Within select UK communities, the health of salmonid populations is of great importance. As discussed at length throughout this thesis one of the major factors affecting salmonid health is parasitic infection, imparting varying degrees of pathogenicity on host populations. Recent technological advances have allowed for the accurate identification of aquatic parasites to monitor the potential presence of highly virulent species, to minimise detrimental impact on fish populations. Ascertaining immunogenic response to parasite infections is one of the key factors in understanding resistance or susceptibility of populations to infection. Recent studies utilizing genomic analysis have allowed for the quantification of immunogenic response cascades post parasitic invasion of host, leading to isolation of potential vaccine targets and important parasite proteins contributing to immunomodulation.

This body of work sought to advance the understanding of host-parasite relationships within brown trout populations. Initial parasite screen was the first molecular prospecting study of the species *Diplostomum* spp. within the United Kingdom, updating previous assessments of species richness and identification of *Diplostomum* spp infecting brown trout in the UK. The second study relating to species identification involved the medically important *Diphyllobothrium dendriticum* infection*;* once again this was the first study to utilize molecular techniques to identify a medically re-emergent parasite infecting UK brown trout. The study area provided an excellent transferable freshwater environment to model the evolutionary properties that maintain immunogenic diversity within salmonid MHC genes. The isolated populations highlighted specificity of parasitic insult and inter population infection diversity in correlation with immunogenetic

diversity, to infer the potential for parasite-driven selection on host populations. With the vast infective success demonstrated by *Diplostomum baeri* within our study area and worldwide, there was a clear need for further investigation into potential molecular factors contributing to infection dynamics of the species. The study was the first to identify transmembrane antigens associated with immune evasion of the parasite and lead to continued infective success.

8.2 Parasite screen of brown trout within Gairloch freshwater system

As a freshwater system, the Gairloch hill lochs provided numerous populations of brown trout existing within varying degrees of isolation. As a study area it shares similar characteristics to freshwater and anadromous habitats of trout worldwide. This makes the region a suitable model system to extrapolate parasite infection dynamics across similar brown trout habitats. The system itself indicated clear differential parasite infection across sub-populations. Within the system there was a clear comparative differential infection present within populations of lower laying western lochs compared to elevated loch sub-populations. Multiple factors can lead to differential parasitism across localities and between freshwater populations of brown trout. One of the factors leading to the presence of cestodes infection within the low lying lochs was increased bird population within the locality, most likely due to a nearby refuse area. Additionally the potential presence of intermediate host *Gasterosteus aculeatus* may have also lead to the presence of tapeworm. A piscivorous diet switch of brown trout to eating stickleback can be a common route of infection for *Diphyllobothrium* and *Eutbothrium.*

Across the loch system eye fluke remained the dominant parasite presence throughout brown trout populations. Although rarely studied within the United Kingdom *Diplostomum* is a ubiquitous parasite infecting brown trout worldwide (Ndeda et al., 2013; Blasco-Costa et al., 2014c; Faltýnková et al., 2014; Selbach et al., 2015b). Its presence within the system indicates a clear piscivourous bird population, maintaining *Diplosotomum*. The presence of snail intermediate host is also important to sustain

infection rate, without performing snail sampling field studies, it is clear that a healthy size snail population exists within the system. Nematode infection was low within the system; *Eustronglyides* is a common parasite of brown trout (Hirshfield et al., 1983) and was only found within two individuals within the entire dataset. The infected individuals are found within larger, deeper lochs. Nematode infective presence within these lochs correlates the presence of *Oligacaete* worms dwelling at greater depths in the water column.

A key outcome of the study highlighted the importance of host presence to support parasite infection within isolated populations. Overall the study provided one important outcome to link to further studies within the body of work, differential parasitism within the Gairloch system, provided the framework for further immune-genetic analysis in correlation with differential infection.

Moving forward, analysis of fauna in freshwater systems would be enhanced by sampling in a longitudinal manner rather than a singular multi-population sampling method. The longitudinal monitoring of freshwater fish populations would allow for the analysis of temporal fluctuation in parasite prevalence and diversity. Temporal variation in parasite prevalence could be due to a multitude of factors including seasonal temperature variations and definitive host movement. Although parasite presence was consistent throughout the system, the actual pathogenic affect of parasite infection is still yet to be determined. A study of parasite infection load alongside growth-rate and age may help in determining the potential detrimental affects that parasite infection is having on trout populations. A study like this would be particularly impactful due to the as yet misunderstood pathogenic-nature of *Diplostomum* infections in freshwater systems, despite being seen as a highly virulent parasite within them.

8.3 Species richness and diversity of *D. baeri* infecting a hill loch system

The infective level of *Diplostomum* spp. was high across the dataset, which prompted further analysis of species identification and genetic diversity. Using *cox1* and ITS genetic markers the species identified as *Diplostomum baeri*. *D. baeri* indicates a continued high prevalence across freshwater salmonids and Percidae worldwide. Its presence within the United Kingdom has been identified before within farmed (Stables and Chappell, 1986b) and wild (Stables and Chappell, 1986a) populations. The study provided the first molecular evidence of *D. baeri* infecting UK brown trout. The discovery of only a single species of *Diplostomum* infecting UK brown trout confirmed by both cox1 and ITS markers shows a deviation from precious Palearctic studies of the species; with 3 species of *Diplostomum* (*D. baeri, D.spathaceum* and *D.pseudospathaceum*) discovered infecting brown trout within Icelandic population of brown trout (Blasco-Costa et al., 2014c) and 4 (*D. baeri, D.mergi, D.spathaceum* and *D.pseudospathaceum)* infecting German populations of brown trout (Georgieva et al., 2013b). The reasoning for single species infection within Gairloch populations may be associated with a "lifecycle bottleneck" within only a single molluscan host infection existing within the loch system maintaining *Diplostomum baeri* infection within the system. Although *D. baeri* was the only species present within the system, it presented as a highly speciated complex with multiple lineages present across the dataset, the *D. baeri* species complex is not only present within this dataset but also across north American (S. A. Locke et al., 2010b) and Icelandic (Blasco-Costa et al., 2014c) *D. baeri* infected localities. The genetic diversity of *D. baeri* is most likely due to definitive host mobility within local geography of Gairloch. The region is host to a large transitory bird presence every summer with annual southern migration of *Gavia artica* and *Gavia stellata*, piscivorous birds with known a known breeding habitat surrounding elevated hill lochs. Through geographic specific haplotypes confirmed the presence of an admixture of Icelandic and mainland European isolates alongside Gairloch specific isolates infecting brown trout in the system. The study provided the first evidence of *D. baeri* infection in the UK backed by mitochondrial and nuclear species identification. It also confirmed the increased diversity of the *D. baeri* species complex present infecting freshwater fish,

isolates deriving from Gairloch place alongside salmonid specific clade in phylogenetic analysis, indicating potential host-specific species complex existing within *D. baeri* a key evolutionary mechanism that may go part way in explaining such a wide spread infection of freshwater fish.

With *Diplostomum baeri* demonstrating comparative increased diversity compared to other species of Diplostomidae in this study and within other publicized studies (Locke *et al*., 2010b; Georgieva *et al*., 2013b; Blasco-Costa *et al.*, 2014c) future genetic analysis of the species in the UK trout should be based around the underlying factors affecting genetic diversity of *D. baeri*. One factor that may affect the genetic diversity within the system may be the continued annual presence of transitory definitive host species. A longitudinal study of *Diplostomum baeri* genetic diversity after seasonal presence of bird species would go part way to explain the potential impact the transitory bird presence has on genetic diversity of *Diplostomum baeri*. With populations of brown trout existing in varying hydrologic ally diverse water bodies in the Gairloch system; hydrological dynamics such as water depth and velocity may play in a role in genetic diversity of *Diplosotmum* infecting intermediate host. The diversity of cercariae infecting freshwater hosts has already been observed within Schistosoma (N'Goran *et al.*, 1997), Echinostoma (Combes *et al.*, 1994) and Diplostomum (Karvonen *et al.*, 2003) where slow moving waters indicate a higher level of trematode infection and an increased level of genetic diversity infecting freshwater hosts. With Gairloch brown trout populations inhabiting lochs and burns with varying hydrological dynamics, a focused sampling effort taking these dynamics into account would allow for the comparison of cercarial diversity within differing water bodies to take place.

8.4 Molecular characterization of re-emergent fish tapeworm infection within Scottish populations of *Salmo trutta*

The identification of fish parasites is of high importance to monitor and control of food born diseases. *Diphyllobothriasis* is considered a current re-emerging fish tape worm disease across Europe, with the spread in popularity of raw fish dishes the disease is seeing a steady increase in across Europe within the last 20 years. Identification of the parasite within freshwater fish is crucial in delineating between medically relevant species and non-human infecting species. The study was based on samples removed from the musculature and intestinal tract of *S.trutta* within select lochs in the Gairloch system revealing. Both *cox1* and ITS genetic markers were used to identify the plerocercoid larval stage as the human pathogenic *D.dendriticum*. Comparisons between the two markers revealed that *cox1* was more accurate in species identification, with ITS marker placing *D.dendriticum* samples alongside *D.ditruem* samples. *D.ditreum* is a non-pathogenic species that also infects salmonids; meaning delineation between the two samples is crucial. Further analysis using this *cox1* marker also proved to be useful in inferring evolutionary insights into the historical movements of Diphyllobothrium. The movement of carnivorous mammal species across Europe is primarily responsible for the trans-placement in reaction to the progression and recession of European ice sheet. Currently there is a significant lack of knowledge on the role the parasite plays in the UK in both medical relevance and its affect on the health of fish populations. Diphyllobothrium is currently considered as a re-emergent food borne disease, molecular data could allow for the effective tracking and monitoring of pathogenic species throughout Europe. Correct genetic marker use to discern potentially pathogenic species of *Diphyllobothrium* is a crucial tool for public health monitoring. The majority of *D.dendriticum* infections currently derives from imported Chilean salmon to Europe (Wicht et al., 2008; Kuchta et al., 2013) the use and implementation of genetic testing on imported whole fish would improve food safety standards of imported fish products.

The discovery of *D.dendriticum* within the system is important due the pathogenic nature of the parasite in humans. Due to its importance, the monitoring and detection of the tapeworm within freshwater bodies should be considered. One such measure may be the development of eDNA methodologies to detect crucial stickleback or copepod primary host presence within a loch. This would allow for the detection of *Diphyllobothrium* to specific localities, minimising impact on fish stock use for the monitoring the parasite presence. With the parasite being one of very few medically relevant parasites within the UK, its continued long term monitoring must be considered, particularly within areas such as Gairloch where angling based tourism is so prevalent.

8.5 Insights into the evolution of salmonid major histocompatibility complex across wild populations of *S.trutta* under differential parasite infection

Wild brown trout frequently exist in isolated lacustrine populations within the Scottish highlands. Brown trout must adapt to the environmental factors to establish sustainable long-lasting population within them. One key factor that populations must address is the response mechanism to parasite infection within local habitat. A key immunogenic response mechanism to parasitic insult is the Major Histocompatibility complex (MHC). As part of the extracellular pathogen immune response the maintained genetic diversity of MHC correlates directly with population fitness. Within the Gairloch system, the varied isolated loch populations allowed for the direct inter population investigation into the host-parasite relationship existing between parasite diversity and MHC response mechanism. The results indicate an adaptive immunogenic response from brown trout populations to habituate to parasite infective presence within immediate environment. Throughout the dataset positive selection exists within the antigen-binding site across brown trout sub-population. Further analysis was based on understanding the underlying reasoning behind positive selection on MHC. Parasite-driven selection indicated a potential role in maintaining MHC diversity within sub-populations. With sub-populations undergoing differential infection potential exhibited greater degree of selective pressure acting on final amino acid motif of ABS. The functional adaptation of

host species to diverse parasite selection highlights the role that parasite infection plays in maintaining genetic diversity of the antigen-binding site in the MHC.

Brown trout populations existing in water bodies with parasitic insult deriving from varied species have adapted immunogenic recognition to counteract the varied infective potential. The result of this study not only has implications for the trans placement of fish for stocking measures but also would mean the stocking of non-native fish species. The stocking of non-native species of salmonids have previously been indicated immunogenic naivety could lead to non-viability of stocked populations (O'Farrell et al., 2013b; C. Monzón-Argüello et al., 2014; Lamaze et al., 2014c). Outcomes from this study agree that fish exposed to a limited parasite milieu may display characteristics of immunogenic naivety. However, in time of stocked populations of salmonids can show adaptive traits to proliferate under natural parasite infection levels. The understanding of the adaptive traits of salmonids affects both conservation practices but also can help ensure that salmonid aquaculture practice can be more sustainable in the future with decreased reliance of pharmalogical measures to prevent parasite infection.

This research has been the first research based on UK MHC diversity of salmonids.
Now with the understanding gained from the work presented here, future work should focus more on the anadromous populations of salmonids that exist within the system. A debate regarding speciation between sea going and residential brown trout has always been at the forefront of speciation studies of salmonids (Gyllensten, 1985; Hindar *et al*., 1991; Jonsson and L'Abee-Lund, 1993; Schluter, 1996). The MHC has been highlighted as a potential gene that contributes to speciation processes in fish (Sato, 2003; Matthews *et al.*, 2010; Malmstrøm *et al.,* 2016; Kaufman, 2018). With ferox, sea and residential morphs of brown trout all existing within the Gairloch system, a cross-morph analysis of MHC diversity may expand the understanding of the MHC's role as a speciation gene within a singular system. In terms of the host-parasite interaction, immunogenetic analysis may have to go beyond the MHC alone. The immune response to parasite infection involves a multitude of potential immune cascades being made up of multiple immune cells. Further studies could concentrate on the adaptation of other parts of this

cascade including; mucosal epithelium (Micallef et al., 2012) or the innate immune system (Alvarez-Pellitero, 2008). Utilizing other factors of the salmonid immune system will give a more complete picture into the true immunogenetic adaptation of salmonid populations to the parasites within its habitat.

8.6 The construction and annotation of *Diplostomum baeri* mitochondrial genome

Diplostomum is currently is one of the most prevalent parasites infecting populations of salmonids worldwide. However information regarding molecular factors underlying the parasites infective success is sorely lacking. The current DNA barcode marker *cox1* has been the primary marker used for the identification and genetic analysis in field studies of *Diplostomum* spp. Within this body of work mitochondrial and nuclear markers have been used to identify the species *D. baeri* infecting Gairloch system trout. With such a close relationship existing between all species of *Diplostomum* there is a clear need to explore alternative genetic markers displaying increased diversity to allow delineation of closely related parasites. *Diplostomum baeri* in-particular exists as a large highly diverse species complex, infecting multiple intermediate fish host species (Blasco-Costa et al., 2014c). The construction of the mitochondrial genome of *D. baeri* revealed a close relationship existing between *D.spathaceum* and *D.pseudospathaceum* in terms of gene order, overall size and phylogenetic relationship across Diplostomidae. Also the comparison of pairwise diversity across the genus highlights potentially more efficacious markers *nad4* and *nad5* from the mitochondrial dataset.

The annotation and construction of the mitochondrial genome of *D. baeri* has allowed for comparison across the previously annotated mitochondrial genomes of *D.spathaceum* and *D.pseudospathaceum*. One way in expanding the work now we have robust methodologies in the construction and annotation of mitochondrial genomes of *Diplostomum* would be to increase the depth and breadth of the study. With genome skimming now being an oft-utilised method in the extraction of hundreds of mitochondrial genomes of mammals (Malé *et al*., 2014), insects (Kocher *et al*., 2015) and pathogens (Denver *et al*., 2016). A similar workflow could be used on *Diplostomum*

metacercariae within trout populations of Gairloch to gain insights into the diversity of the species in a more robust barcoding approach, instead of relying on the frequently used *cox1*. It would also increase the knowledge of potential effectiveness in single mitochondrial marker use in assessing closely related species of *Diplostomum*.

8.7 Genomic mining of intracellular vaccine targets in the fish eye fluke *Diplsotomum baeri*

The infective success of *D. baeri* exists as a complex relationship between host and the parasite. During the blood stage of infection the parasite is under a barrage of immune cells in response to invasion, with a varied cocktail of adaptive and innate immune response cascades utilized to defend the host. However, with such a high level of prevalence within the Gairloch system *D. baeri* must possess certain immunomodulatory capabilities. Across other species of freshwater trematode sharing similar complex freshwater life cycles, the transmembrane proteins tetraspanins (Huang et al., 2005) and Venom allergen-like antigens (Chalmers and Hoffmann, 2012) are presented on the tegument outer layer to provide an immunomodulatory characteristics ensuring parasite avoids host immune system recognition. Using genomic annotation and proteonomic modelling these antigens were discovered within *D. baeri* also. The proteins could be implicated as a factor in the infective success that is demonstrated by *D. baeri*, also indicating a conserved immunomodulatory response within freshwater Digenea. *D. baeri* being so ubiquitous within freshwater fish populations, isolating potentially important antigenic proteins could provide vaccine targets for further evaluation to be used to control the disease within wild populations and the farmed trout setting.

The discovery of the tegumental antigens in this study is an important discovery. However the actual investigation into the importance of these antigens in *Diplosotmum baeri* has yet to be confirmed. For the antigens to be treated as potential aquaculturally important vaccine targets has yet to be determined. To do so would require gene knockdown studies, antigen titration studies and protein expression studies within the migrating parasite, which are all a long way off in terms of vaccine development. The

tetraspanin's in particular represent an interesting marker to be examined in terms of population genetics across parasite populations. Measuring the genetic diversity of the tetraspanin antigens across different populations has been done before within *S.mansoni* populations (Sealey *et al.,* 2013). Now that antigenic sequences have been mined from the genome, region specific oligos can be designed to study the antigens on a population level. The study would compare and contrast populations that see increased levels of *Diplostomum* infection to gain insight into the impact that antigen diversity would have on the virulence of *Diplostomum* within a freshwater fish population. With the annotation of epitope binding domains on tetraspanins, regions could be selected that interact directly with the host-immune system such as the extracellular loop, increasing the likelihood that regions of the gene studied could directly impact the affectiveness of immunoevasion within the host.

8.8 Limitations

The work was performed using a sample archive, with the sampling areas being extensive and varied in terms of geography and hydrology. However with all wild fish studies the conclusions you draw from the study could become more concise with both an increased sample site number and increased population number. With the study attempting to answer population wide questions the increase in the size of the sample populations could have provided stronger statistical support than currently obtained. The other limitation is the cost of Next-generation sequencing, with the procedure being quite costly the study could only manage enough sequencing runs for one complete *D. baeri* genome. With the further amplification of multiple *D. baeri* genomes from different localities in the study area, could have provided tegumental protein comparisons of different locations of *Diplostomum* to allowing deeper understanding of the roles played by TSP and VAL antigenic proteins in parasite infection.

9 Bibliography

Abath, F.G.C., Werkhauser, R.C., 1996. The tegument of Schistosoma mansoni: functional and immunological features. Parasite Immunol. 18, 15–20. https://doi.org/10.1046/j.1365-3024.1996.d01-6.x

Akira, S., Uematsu, S., Takeuchi, O., 2006. Pathogen Recognition and Innate Immunity. Cell 124, 783–801. https://doi.org/10.1016/j.cell.2006.02.015

Alba, A., Hernández, H.M., Marcet, R., Gil, A.L., Vázquez, A.A., Figueredo, M., Sánchez, J., Garay, H.E., Sarracent, J., 2014. Exploring the antigenic features of Fasciola hepatica rediae (Trematoda: Digenea) through the evaluation of different antigenic candidates for further monoclonal antibody generation. Parasitol. Res. 113, 3185–3193. https://doi.org/10.1007/s00436-014-3981-y

Alvarez-Pellitero, P., 2008. Fish immunity and parasite infections: from innate immunity to immunoprophylactic prospects. Vet. Immunol. Immunopathol. 126, 171–198. https://doi.org/10.1016/j.vetimm.2008.07.013

Araujo, A., Reinhard, K.J., Ferreira, L.F., Gardner, S.L., 2008. Parasites as probes for prehistoric human migrations? Trends Parasitol. 24, 112–115. https://doi.org/10.1016/j.pt.2007.11.007

Ardelli, B.F., Woo, P.T.K., 2001. Therapeutic and prophylactic effects of isometamidium chloride (Samorin) against the hemoflagellate Cryptobia salmositica in chinook salmon (Oncorhynchus tshawytscha) and the effects of the drug on uninfected rainbow trout (O. mykiss). Parasitol. Res. 87, 18–26. https://doi.org/10.1007/s004360000294

Arneberg, P., 2002. Host population density and body mass as determinants of species richness in parasite communities: comparative analyses of directly transmitted nematodes of mammals. Ecography 25, 88–94. https://doi.org/10.1034/j.1600-0587.2002.250110.x

Bakal, R.S., Hickson, B.H., Gilger, B.C., Levy, M.G., Flowers, J.R., Khoo, L., 2005. Surgical removal of cataracts due to diplostomum species in gulf sturgeon (acipenser oxyrinchus desotoi). J. Zoo Wildl. Med. 36, 504–508. https://doi.org/10.1638/04-044.1

Barber, I., Hoare, D., Krause, J., 2000. Effects of parasites on fish behaviour: a review and evolutionary perspective 36.

Bass, D., Stentiford, G.D., Littlewood, D.T.J., Hartikainen, H., 2015. Diverse Applications of Environmental DNA Methods in Parasitology. Trends Parasitol. 31, 499–513. https://doi.org/10.1016/j.pt.2015.06.013

Behrmann-Godel, J., 2013. Parasite identification, succession and infection pathways in perch fry (Perca fluviatilis): new insights through a combined morphological and genetic approach. Parasitology 140, 509–20. https://doi.org/10.1017/S0031182012001989

Beldomenico, P.M., Begon, M., 2010. Disease spread, susceptibility and infection intensity: vicious circles? Trends Ecol. Evol. 25, 21–27. https://doi.org/10.1016/j.tree.2009.06.015

Bernatchez, Louis, 2001. The evolutionary history of brown trout (salmo trutta l.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. Evolution 55, 351–379. https://doi.org/10.1111/j.0014-3820.2001.tb01300.x

Bernatchez, L, 2001. The evolutionary history of brown trout (Salmo trutta L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. Evol. Int. J. Org. Evol. 55, 351–79. https://doi.org/10.2307/2640757

Bernatchez, L., Landry, C., 2003. MHC studies in nonmodel vertebrates : what have we learned about natural selection in 15 years ? 16, 363–377.

Betterton, C., 1974. Studies on the host specificity of the eyefluke, Diplostomum spathaceum, in brown and rainbow trout. Parasitology 69, 11. https://doi.org/10.1017/S003118200004614X

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42, W252–W258. https://doi.org/10.1093/nar/gku340

Bjørn, P.A., Finstad, B., Kristoffersen, R., 2001. Salmon lice infection of wild sea trout and Arctic char in marine and freshwaters: The effects of salmon farms. Aquac. Res. 32, 947–962. https://doi.org/10.1046/j.1365-2109.2001.00627.x

Blasco-Costa, Isabel, Cutmore, S.C., Miller, T.L., Nolan, M.J., 2016. Molecular approaches to trematode systematics: 'best practice' and implications for future study. Syst. Parasitol. 93, 295–306. https://doi.org/10.1007/s11230-016-9631-2

Blasco-Costa, I., Faltýnková, A., Georgieva, S., Skírnisson, K., Scholz, T., Kostadinova, A., 2014a. Fish pathogens near the Arctic Circle: Molecular, morphological and ecological evidence for unexpected diversity of diplostomum (Digenea: Diplostomidae) in Iceland. Int. J. Parasitol. 44, 703–715. https://doi.org/10.1016/j.ijpara.2014.04.009

Blasco-Costa, I., Faltýnková, A., Georgieva, S., Skírnisson, K., Scholz, T., Kostadinova, A., 2014b. Fish pathogens near the Arctic Circle: molecular, morphological and ecological evidence for unexpected diversity of Diplostomum (Digenea: diplostomidae) in Iceland. Int. J. Parasitol. 44, 703–15. https://doi.org/10.1016/j.ijpara.2014.04.009

Blasco-Costa, I., Faltýnková, A., Georgieva, S., Skírnisson, K., Scholz, T., Kostadinova, A., 2014c. Fish pathogens near the Arctic Circle: molecular, morphological and ecological evidence for unexpected diversity of Diplostomum (Digenea: diplostomidae) in Iceland. Int. J. Parasitol. 44, 703–715. https://doi.org/10.1016/j.ijpara.2014.04.009

Blasco-Costa, I., Poulin, R., Presswell, B., 2016. Morphological description and molecular analyses of Tylodelphys sp. (Trematoda: Diplostomidae) newly recorded from the freshwater fish Gobiomorphus cotidianus (common bully) in New Zealand. J. Helminthol. 1–14. https://doi.org/10.1017/S0022149X16000298

Borghans, J.A.M., Beltman, J.B., De Boer, R.J., 2004. MHC polymorphism under host-pathogen coevolution. Immunogenetics 55, 732–739. https://doi.org/10.1007/s00251-003-0630-5

Borgstrøm, R., Trømborg, J., Haugen, T.O., Rosseland, B.O., 2017. Plerocercoids of the cestode *Diphyllobothrium ditremum* in brown trout *Salmo trutta* : substantial increase in infection after establishment of European minnow *Phoxinus phoxinus*: *diphyllobothrium* , in *s. trutta* and *p. phoxinus*. J. Fish Biol. 91, 912–927. https://doi.org/10.1111/jfb.13391

Boyce, N.P., Yamada, S.B., 1977. Effects of a Parasite, *Eubothrium salvelini* (Cestoda: Pseudophyllidea), on the Resistance of Juvenile Sockeye Salmon, *Oncorhynchus nerka* , to Zinc. J. Fish. Res. Board Can. 34, 706–709. https://doi.org/10.1139/f77-110

Brabec, J., Kostadinova, A., Scholz, T., Littlewood, D.T.J., 2015a. Complete mitochondrial genomes and nuclear ribosomal RNA operons of two species of Diplostomum (Platyhelminthes: Trematoda): a molecular resource for taxonomy and molecular epidemiology of important fish pathogens. Parasit. Vectors 8. https://doi.org/10.1186/s13071-015-0949-4

Brabec, J., Kostadinova, A., Scholz, T., Littlewood, D.T.J., 2015b. Complete mitochondrial genomes and nuclear ribosomal RNA operons of two species of Diplostomum (Platyhelminthes: Trematoda): a molecular resource for taxonomy and molecular epidemiology of important fish pathogens. Parasit. Vectors 8, 336. https://doi.org/10.1186/s13071-015-0949-4

Bracamonte, S.E., Baltazar-Soares, M., Eizaguirre, C., 2015. Characterization of MHC class II genes in the critically endangered European eel (Anguilla anguilla). Conserv. Genet. Resour. 7, 859–870. https://doi.org/10.1007/s12686-015-0501-z

Brant, S.V., Morgan, J.A.T., Mkoji, G.M., Snyder, S.D., Rajapakse, R.P.V.J., Loker, E.S., 2006. An approach to revealing blood fluke life cycles, taxonomy, and diversity: provision of key reference data including dna sequence from single life cycle stages. J. Parasitol. 92, 77–88. https://doi.org/10.1645/GE-3515.1

Buchmann, K., Bresciani, J., 1997. Parasitic infections in pond-reared rainbow trout Oncorhynchus mykiss in Denmark. Dis. Aquat. Organ. 28, 125–138. https://doi.org/10.3354/dao028125

Buchmann, K., Uldal, A., 1997. Gyrodactylus derjavini infections in four salmonids : comparative host susceptibility and site selection of parasites.

Bush, S.E., Villa, S.M., Altuna, J.C., Johnson, K.P., Shapiro, M.D., Clayton, D.H., 2018. Host defense triggers rapid adaptive radiation in experimentally evolving parasites. bioRxiv. https://doi.org/10.1101/420380

Byrne, C.J., Holland, C. V, Poole, R., Kennedy, C.R., 2002. Comparison of the macroparasite communities of wild and stocked brown trout (Salmo trutta L.) in the west of Ireland. Parasitology 124, 435–445. https://doi.org/10.1017/S0031182001001330

Campos, J.L., Posada, D., Morán, P., 2006. Genetic variation at MHC, mitochondrial and microsatellite loci in isolated populations of Brown trout (Salmo trutta). Conserv. Genet. 7, 515–530. https://doi.org/10.1007/s10592-005-9063-z

Chalmers, I.W., Hoffmann, K.F., 2012. Platyhelminth Venom Allergen-Like (VAL) proteins: revealing structural diversity, class-specific features and biological

associations across the phylum. Parasitology 139, 1231–1245. https://doi.org/10.1017/S0031182012000704
Chalmers, I.W., McArdle, A.J., Coulson, R.M., Wagner, M.A., Schmid, R., Hirai, H., Hoffmann, K.F., 2008a. Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the Schistosoma mansoni venom allergen-like (SmVAL) gene family. BMC Genomics 9, 89. https://doi.org/10.1186/1471-2164-9-89
Chalmers, I.W., McArdle, A.J., Coulson, R.M., Wagner, M.A., Schmid, R., Hirai, H., Hoffmann, K.F., 2008b. Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the Schistosoma mansoni venom allergen-like (SmVAL) gene family. BMC Genomics 9, 89. https://doi.org/10.1186/1471-2164-9-89
Ciborowski, K., Jordan, W.C., Garcia de Leaniz, C., Consuegra, S., 2017. Temporal and spatial instability in neutral and adaptive (MHC) genetic variation in marginal salmon populations. Sci. Rep. 7, 42416. https://doi.org/10.1038/srep42416
Coceres, V.M., Alonso, A.M., Nievas, Y.R., Midlej, V., Frontera, L., Benchimol, M., Johnson, P.J., de Miguel, N., 2015. The C-terminal tail of tetraspanin proteins regulates their intracellular distribution in the parasite *T richomonas vaginalis*: C-terminal tail of tetraspanin proteins. Cell. Microbiol. 17, 1217–1229. https://doi.org/10.1111/cmi.12431
Cohen, S., 2002. Strong Positive Selection and Habitat-Specific Amino Acid Substitution Patterns in Mhc from an Estuarine Fish Under Intense Pollution Stress. Mol. Biol. Evol. 19, 1870–1880. https://doi.org/10.1093/oxfordjournals.molbev.a004011
Combes, C., Fournier, A., Moné, H., Théron, A., 1994. Behaviours in trematode cercariae that enhance parasite transmission: patterns and processes. Parasitology 109, S3–S13. https://doi.org/10.1017/S0031182000085048
Conejeros, P., Phan, A., Power, M., Alekseyev, S., O'Connell, M., Dempson, B., Dixon, B., 2008. MH class IIα polymorphism in local and global adaptation of Arctic charr (Salvelinus alpinus L.). Immunogenetics 60, 325–337. https://doi.org/10.1007/s00251-008-0290-6
Consuegra, S., Garcia de Leaniz, C., 2008. MHC-mediated mate choice increases parasite resistance in salmon. Proc. Biol. Sci. 275, 1397–403. https://doi.org/10.1098/rspb.2008.0066
Cooper, M.D., Alder, M.N., 2006. The Evolution of Adaptive Immune Systems. Cell 124, 815–822. https://doi.org/10.1016/j.cell.2006.02.001
Cousyn, C., De Meester, L., Colbourne, J.K., Brendonck, L., Verschuren, D., Volckaert, F., 2001. Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. Proc. Natl. Acad. Sci. U. S. A. 98, 6256–6260. https://doi.org/10.1073/pnas.111606798
Cowman, A.F., Crabb, B.S., 2006. Invasion of Red Blood Cells by Malaria Parasites. Cell 124, 755–766. https://doi.org/10.1016/j.cell.2006.02.006
Cribb, T.H., Pearson, J.C., publication, A., n.d. Neodiplostomum spratti n. sp. (Digenea: Dipiostomidae) from Antechinus spp. (Marsupialia: Dasyuridae) in Australia, with notes on other diplostomids from Australian mammals 11.

Criscione, C.D., Blouin, M.S., 2005. Eleven polymorphic microsatellite loci for the salmonid trematode Plagioporus shawi. Mol. Ecol. Notes 5, 562–564. https://doi.org/10.1111/j.1471-8286.2005.00991.x
Criscione, C.D., Blouin, M.S., 2004. Life cycles shape parasite evolution: comparative population genetics of salmon trematodes. Evolution 58, 198–202. https://doi.org/10.1111/j.0014-3820.2004.tb01587.x
Croisetière, S., Tarte, P.D., Bernatchez, L., Belhumeur, P., 2008. Identification of MHC class IIβ resistance/susceptibility alleles to Aeromonas salmonicida in brook charr (Salvelinus fontinalis). Mol. Immunol. 45, 3107–3116. https://doi.org/10.1016/j.molimm.2008.03.007
Crossin, G.T., Hinch, S.G., Farrell, A.P., Higgs, D.A., Lotto, A.G., Oakes, J.D., Healey, M.C., 2004. Energetics and morphology of sockeye salmon: Effects of upriver migratory distance and elevation. J. Fish Biol. 65, 788–810. https://doi.org/10.1111/j.1095-8649.2004.00486.x
Cucherousset, J., Ombredane, D., Charles, K., Marchand, F., Baglinière, J.-L., 2005. A continuum of life history tactics in a brown trout (*Salmo trutta*) population. Can. J. Fish. Aquat. Sci. 62, 1600–1610. https://doi.org/10.1139/f05-057
Cuesta, A., Ángeles Esteban, M., Meseguer, J., 2006. Cloning, distribution and up-regulation of the teleost fish MHC class II alpha suggests a role for granulocytes as antigen-presenting cells. Mol. Immunol. 43, 1275–1285. https://doi.org/10.1016/j.molimm.2005.07.004
Curwen, R.S., Ashton, P.D., Sundaralingam, S., Wilson, R.A., n.d. Identification of Novel Proteases and Immunomodulators in the Secretions of Schistosome Cercariae That Facilitate Host Entry*□S 10.
Cutrera, A.P., Zenuto, R.R., Lacey, E.A., 2014. Interpopulation differences in parasite load and variable selective pressures on MHC genes in Ctenomys talarum. J. Mammal. 95, 679–695. https://doi.org/10.1644/13-MAMM-A-120
Dalton, J.P., Brindley, P.J., Knox, D.P., Brady, C.P., Hotez, P.J., Donnelly, S., O'Neill, S.M., Mulcahy, G., Loukas, A., 2003. Helminth vaccines: from mining genomic information for vaccine targets to systems used for protein expression. Int. J. Parasitol. 33, 621–640. https://doi.org/10.1016/S0020-7519(03)00057-2
Davies, G., Stear, M.J., Benothman, M., Abuagob, O., Kerr, A., Mitchell, S., Bishop, S.C., 2006. Quantitative trait loci associated with parasitic infection in Scottish blackface sheep. Heredity 96, 252–258. https://doi.org/10.1038/sj.hdy.6800788
De Boer, R.J., Borghans, J.A.M., van Boven, M., Keşmir, C., Weissing, F.J., 2004. Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. Immunogenetics 55, 725–731. https://doi.org/10.1007/s00251-003-0629-y
de León, G.P.-P., Choudhury, A., 2010. Parasite Inventories and DNA-based Taxonomy: Lessons from Helminths of Freshwater Fishes in a Megadiverse Country. J. Parasitol. 96, 236–244. https://doi.org/10.1645/GE-2239.1
Denver, D.R., Brown, A.M.V., Howe, D.K., Peetz, A.B., Zasada, I.A., 2016. Genome Skimming: A Rapid Approach to Gaining Diverse Biological Insights into Multicellular Pathogens. PLOS Pathog. 12, e1005713. https://doi.org/10.1371/journal.ppat.1005713
Dezfuli, B.S., Giari, L., Biaggi, S.D., Poulin, R., n.d. Associations and interactions among intestinal helminths of the brown trout, Salmo trutta, in northern Italy 6.

Dick, T.A., Papst, M.H., Paul, H.C., 1987. Rainbow trout (salmo gairdneri) stocking and contracaecum spp. J. Wildl. Dis. 23, 242–247. https://doi.org/10.7589/0090-3558-23.2.242
Dickson, J.H., 1977. Scottish woodlands; their ancient past and precarious future 46, 155–165.
Dionne, M., Miller, K.M., Dodson, J.J., Bernatchez, L., 2009. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. Philos. Trans. R. Soc. B Biol. Sci. 364, 1555–1565. https://doi.org/10.1098/rstb.2009.0011
Dionne, M., Miller, K.M., Dodson, J.J., Caron, F., Bernatchez, L., 2007. Clinal variation in mhc diversity with temperature: evidence for the role of host pathogen interaction on local adaptation in atlantic salmon. Evolution 61, 2154–2164. https://doi.org/10.1111/j.1558-5646.2007.00178.x
Dorucu, M., Crompton, D.W.T., Huntingford, F.A., Walters, D.E., 1995. The ecology of endoparasitic helminth infections of brown trout (Salmo trutta) and rainbow trout (Oncorhynchus mykiss) in Scotland. Folia Parasitol. (Praha).
Duguid, J.B., Sheppard, E.M., 1944. A diphyllobothrium epidemic in trout. J. Pathol. Bacteriol. 56, 73–80. https://doi.org/10.1002/path.1700560110
Dunn, P.O., Bollmer, J.L., Freeman-Gallant, C.R., Whittingham, L.A., 2013. MHC variation is related to a sexually selected ornament, survival, and parasite resistance in common yellowthroats: mhc-based mate choice in a warbler. Evolution 67, 679–687. https://doi.org/10.1111/j.1558-5646.2012.01799.x
Eizaguirre, C., Lenz, T.L., Kalbe, M., Milinski, M., 2012. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. Nat. Commun. 3, 621. https://doi.org/10.1038/ncomms1632
Eizaguirre, C., Lenz, T.L., Sommerfeld, R.D., Harrod, C., Kalbe, M., Milinski, M., 2011. Parasite diversity, patterns of MHC II variation and olfactory based mate choice in diverging three-spined stickleback ecotypes. Evol. Ecol. 25, 605–622. https://doi.org/10.1007/s10682-010-9424-z
Ejsmond, M.J., Radwan, J., 2011. MHC diversity in bottlenecked populations: a simulation model. Conserv. Genet. 12, 129–137. https://doi.org/10.1007/s10592-009-9998-6
Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M., Pieper, U., Sali, A., 2006. Comparative Protein Structure Modeling Using Modeller. Curr. Protoc. Bioinforma. 15, 5.6.1-5.6.30. https://doi.org/10.1002/0471250953.bi0506s15
Evans, M.L., Neff, B.D., 2009. Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). Mol. Ecol. 18, 4716–4729. https://doi.org/10.1111/j.1365-294X.2009.04374.x
Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evol. Bioinforma. 1, 117693430500100. https://doi.org/10.1177/117693430500100003
Faltýnková, A., Georgieva, S., Kostadinova, A., Blasco-Costa, I., Scholz, T., Skírnisson, K., 2014. Diplostomum von Nordmann, 1832 (Digenea: Diplostomidae) in the sub-Arctic: descriptions of the larval stages of six species discovered recently in Iceland. Syst. Parasitol. 89, 195–213. https://doi.org/10.1007/s11230-014-9517-0

Faltýnková, A., Sures, B., Kostadinova, A., 2016. Biodiversity of trematodes in their intermediate mollusc and fish hosts in the freshwater ecosystems of Europe. Syst. Parasitol. 93, 283–293. https://doi.org/10.1007/s11230-016-9627-y

Fang, F.C., Billman, Z.P., Wallis, C.K., Abbott, A.N., Olson, J.C., Dhanireddy, S., Murphy, S.C., 2015. Human Diphyllobothrium nihonkaiense Infection in Washington State. J. Clin. Microbiol. 53, 1355–1357. https://doi.org/10.1128/JCM.00065-15

Ferguson, A., 1989. Genetic differences among brown trout, Salmo trutta, stocks and their importance for the conservation and management of the species. Freshw. Biol. 21, 35–46. https://doi.org/10.1111/j.1365-2427.1989.tb01346.x

Fermino, B.R., Paiva, F., Soares, P., Tavares, L.E.R., Viola, L.B., Ferreira, R.C., Botero-Arias, R., de-Paula, C.D., Campaner, M., Takata, C.S.A., Teixeira, M.M.G., Camargo, E.P., 2015. Field and experimental evidence of a new caiman trypanosome species closely phylogenetically related to fish trypanosomes and transmitted by leeches. Int. J. Parasitol. Parasites Wildl. 4, 368–378. https://doi.org/10.1016/j.ijppaw.2015.10.005

Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharakhov, I.V., Jiang, X., Hall, A.B., Catteruccia, F., Kakani, E., Mitchell, S.N., Wu, Y.-C., Smith, H.A., Love, R.R., Lawniczak, M.K., Slotman, M.A., Emrich, S.J., Hahn, M.W., Besansky, N.J., 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science 347, 1258524–1258524. https://doi.org/10.1126/science.1258524

Forsberg, L. a., Dannewitz, J., Petersson, E., Grahn, M., 2007. Influence of genetic dissimilarity in the reproductive success and mate choice of brown trout - Females fishing for optimal MHC dissimilarity. J. Evol. Biol. 20, 1859–1869. https://doi.org/10.1111/j.1420-9101.2007.01380.x

Fraser, D. J., Weir, L.K., Bernatchez, L., Hansen, M.M., Taylor, E.B., 2011. Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. Heredity 106, 404–420. https://doi.org/10.1038/hdy.2010.167

Fraser, D J, Weir, L.K., Bernatchez, L., Hansen, M.M., Taylor, E.B., 2011. Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. Heredity 106, 404–420. https://doi.org/10.1038/hdy.2010.167

Fraser, P.G., 1960. The Occurrence of Diphyllobothrium in Trout, with Special Reference to an Outbreak in the West of England. J. Helminthol. 34, 59. https://doi.org/10.1017/S0022149X00020356

Frimeth, J.P., 1987. A survey of the parasites of nonanadromous and anadromous brook charr (*Salvelinus fontinalis*) in the Tabusintac River, New Brunswick, Canada. Can. J. Zool. 65, 1354–1362. https://doi.org/10.1139/z87-215

Galazzo, D.E., Dayanandan, S., Marcogliese, D.J., McLaughlin, J.D., 2002. Molecular systematics of some North American species of Diplostomum (Digenea) based on rDNA-sequence data and comparisons with European congeners. Can. J. Zool. Ott. 80, 2207–2217.

Garcia-España, A., Chung, P.-J., Sarkar, I.N., Stiner, E., Sun, T.-T., DeSalle, R., 2008. Appearance of new tetraspanin genes during vertebrate evolution. Genomics 91, 326–334. https://doi.org/10.1016/j.ygeno.2007.12.005

Gendron, A.D., Marcogliese, D.J., 2017. Enigmatic decline of a common fish parasite (Diplostomum spp.) in the St. Lawrence River: Evidence for a dilution effect induced by the invasive round goby. Int. J. Parasitol. Parasites Wildl. 6, 402–411. https://doi.org/10.1016/j.ijppaw.2017.04.002

Georgieva, S., Soldánová, M., Pérez-del-Olmo, A., Dangel, D.R., Sitko, J., Sures, B., Kostadinova, A., 2013a. Molecular prospecting for European Diplostomum (Digenea: Diplostomidae) reveals cryptic diversity. Int. J. Parasitol. 43, 57–72. https://doi.org/10.1016/j.ijpara.2012.10.019

Georgieva, S., Soldánová, M., Pérez-del-Olmo, A., Dangel, D.R., Sitko, J., Sures, B., Kostadinova, A., 2013b. Molecular prospecting for European Diplostomum (Digenea: Diplostomidae) reveals cryptic diversity. Int. J. Parasitol. 43, 57–72. https://doi.org/10.1016/j.ijpara.2012.10.019

Georgieva, S., Soldánová, M., Pérez-Del-Olmo, A., Dangel, D.R., Sitko, J., Sures, B., Kostadinova, A., 2013c. Molecular prospecting for European Diplostomum (Digenea: Diplostomidae) reveals cryptic diversity. Int. J. Parasitol. 43, 57–72. https://doi.org/10.1016/j.ijpara.2012.10.019

Gharrett, A.J., Joyce, J., Smoker, W.W., 2013. Fine-scale temporal adaptation within a salmonid population: mechanism and consequences. Mol. Ecol. 22, 4457–4469. https://doi.org/10.1111/mec.12400

Gíslason, G.M., Steingrímsson, S.Ó., 2004. Seasonal and spatial variation in the diet of brown trout (Salmo trutta L.) in the subarctic River Laxá, north-east Iceland. Aquat. Ecol. 38, 263–270. https://doi.org/10.1023/B:AECO.0000032052.04874.fb

Gobert, G.N., Stenzel, D.J., McManus, D.P., Jones, M.K., 2003. The ultrastructural architecture of the adult Schistosoma japonicum tegument. Int. J. Parasitol. 33, 1561–1575. https://doi.org/10.1016/S0020-7519(03)00255-8

Godel, C., Kumar, S., Koutsovoulos, G., Ludin, P., Nilsson, D., Comandatore, F., Wrobel, N., Thompson, M., Schmid, C.D., Goto, S., Bringaud, F., Wolstenholme, A., Bandi, C., Epe, C., Kaminsky, R., Blaxter, M., Mäser, P., 2012. The genome of the heartworm, Dirofilaria immitis, reveals drug and vaccine targets. FASEB J. 26, 4650–4661. https://doi.org/10.1096/fj.12-205096

Gomez-Uchida, D., Knight, T.W., Ruzzante, D.E., 2009. Interaction of landscape and life history attributes on genetic diversity, neutral divergence and gene flow in a pristine community of salmonids. Mol. Ecol. 18, 4854–4869. https://doi.org/10.1111/j.1365-294X.2009.04409.x

GoüY De Bellocq, J., Charbonnel, N., Morand, S., 2008. Coevolutionary relationship between helminth diversity and MHC class II polymorphism in rodents. J. Evol. Biol. 21, 1144–1150. https://doi.org/10.1111/j.1420-9101.2008.01538.x

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebø, S., Stet, R.J.M., 2003. MHC polymorphism and disease resistance in Atlantic salmon (Salmo salar); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics 55, 210–219. https://doi.org/10.1007/s00251-003-0567-8

Grobbelaar, A., van As, L.L., van As, J.G., Butler, H.J., 2015. Pathology of eyes and brain of fish infected with diplostomids, southern Africa. Afr. Zool. 50, 181–186. https://doi.org/10.1080/15627020.2015.1055701

Guo, A.-J., Liu, K., Gong, W., Luo, X.-N., Yan, H.-B., Zhao, S.-B., Hu, S.-N., Jia, W.-Z., 2012. Molecular identification of Diphyllobothrium latum and a brief review of diphyllobothriosis in China. Acta Parasitol. 57. https://doi.org/10.2478/s11686-012-0036-3

Gyllensten, U., 1985. The genetic structure of fish: differences in the intraspecific distribution of biochemical genetic variation between marine, anadromous, and freshwater species. J. Fish Biol. 26, 691–699. https://doi.org/10.1111/j.1095-8649.1985.tb04309.x

Haas, W., Wulff, C., Grabe, K., Meyer, V., Haeberlein, S., 2007. Navigation within host tissues: cues for orientation of Diplostomum spathaceum (Trematoda) in fish towards veins, head and eye. Parasitology 134, 1013. https://doi.org/10.1017/S0031182007002430

Hajibabaei, M., Singer, G. a C., Hebert, P.D.N., Hickey, D. a, 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. TIG 23, 167–72. https://doi.org/10.1016/j.tig.2007.02.001

Hakalahti, T., Karvonen, a, Valtonen, E.T., 2006. Climate warming and disease risks in temperate regions--Argulus coregoni and Diplostomum spathaceum as case studies. J. Helminthol. 80, 93–98. https://doi.org/10.1079/joh2006351

Hamilton, C.M., Dowling, D.J., Loscher, C.E., Morphew, R.M., Brophy, P.M., O'Neill, S.M., 2009. The Fasciola hepatica Tegumental Antigen Suppresses Dendritic Cell Maturation and Function. Infect. Immun. 77, 2488–2498. https://doi.org/10.1128/IAI.00919-08

Hansell, E., Braschi, S., Medzihradszky, K.F., Sajid, M., Debnath, M., Ingram, J., Lim, K.C., McKerrow, J.H., 2008. Proteomic Analysis of Skin Invasion by Blood Fluke Larvae. PLoS Negl. Trop. Dis. 2, e262. https://doi.org/10.1371/journal.pntd.0000262

Hansen, M.M., Ruzzante, D.E., Nielsen, E.E., Bekkevold, D., Mensberg, K.-L.D., 2002. Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (Salmo trutta) populations: GENETIC VARIATION IN ANADROMOUS BROWN TROUT POPULATIONS. Mol. Ecol. 11, 2523–2535. https://doi.org/10.1046/j.1365-294X.2002.01634.x

Hansson, B., Westerberg, L., 2002. On the correlation between heterozygosity and fitness in natural populations. Mol. Ecol. 11, 2467–2474. https://doi.org/10.1046/j.1365-294X.2002.01644.x

Hanzelová, V., Scholz, T., Gerdeaux, D., Kuchta, R., 2002. A comparative study of Eubothrium salvelini and E. crassum (Cestoda: Pseudophyllidea) parasites of Arctic charr and brown trout in alpine lakes, in: Magnan, P., Audet, C., Glémet, H., Legault, M., Rodríguez, M.A., Taylor, E.B. (Eds.), Ecology, Behaviour and Conservation of the Charrs, Genus Salvelinus. Springer Netherlands, Dordrecht, pp. 245–256. https://doi.org/10.1007/978-94-017-1352-8_22

Hartvigsen, R., Kennedy, C., 1993. Patterns in the composition and Richness of helmith communities in brown trout, Salmo trutta, in a group pf reservoirs. J. Fish Biol.

Hedrick, P.W., Kim, T.J., Parker, K.M., 2001. Parasite resistance and genetic variation in the endangered Gila topminnow. Anim. Conserv. 4, 103–109. https://doi.org/10.1017/S1367943001001135

Hemler, M.E., 2003. Tetraspanin Proteins Mediate Cellular Penetration, Invasion, and Fusion Events and Define a Novel Type of Membrane Microdomain. Annu. Rev. Cell Dev. Biol. 19, 397–422. https://doi.org/10.1146/annurev.cellbio.19.111301.153609
Hemler, M.E., 2001. Specific tetraspanin functions. J. Cell Biol. 155, 1103–1108. https://doi.org/10.1083/jcb.200108061
Hendrich, L., Pons, J., Ribera, I., Balke, M., 2010. Mitochondrial Cox1 Sequence Data Reliably Uncover Patterns of Insect Diversity But Suffer from High Lineage-Idiosyncratic Error Rates. PLoS ONE 5, e14448. https://doi.org/10.1371/journal.pone.0014448
Hewitson, J.P., Harcus, Y., Murray, J., van Agtmaal, M., Filbey, K.J., Grainger, J.R., Bridgett, S., Blaxter, M.L., Ashton, P.D., Ashford, D., Curwen, R.S., Wilson, R.A., Dowle, A.A., Maizels, R.M., 2011. Proteomic analysis of secretory products from the model gastrointestinal nematode Heligmosomoides polygyrus reveals dominance of Venom Allergen-Like (VAL) proteins. J. Proteomics 74, 1573–1594. https://doi.org/10.1016/j.jprot.2011.06.002
Hindar, K., Jonsson, B., Ryman, N., Ståhl, G., 1991. Genetic relationships among landlocked, resident and anadromous Brown Trout, Salmo trutta L. Heredity 66, 83–91. https://doi.org/10.1038/hdy.1991.11
Hindar, K., Tufto, J., Sattem, L., Balstad, T., 2004. Conservation of genetic variation in harvested salmon populations. ICES J. Mar. Sci. 61, 1389–1397. https://doi.org/10.1016/j.icesjms.2004.08.011
Hirshfield, M.F., Morin, R.P., Hepner, D.J., 1983. Increased prevalence of larval Eustronglylides (Nematoda) in the mummichog, Fundulus heteroclitus (L.), from the discharge canal of a power plant in the Chesapeake Bay. J. Fish Biol. 23, 135–142. https://doi.org/10.1111/j.1095-8649.1983.tb02888.x
Huang, S., Yuan, S., Dong, M., Su, J., Yu, C., Shen, Y., Xie, X., Yu, Y., Yu, X., Chen, S., Zhang, S., Pontarotti, P., Xu, A., 2005. The phylogenetic analysis of tetraspanins projects the evolution of cell–cell interactions from unicellular to multicellular organisms. Genomics 86, 674–684. https://doi.org/10.1016/j.ygeno.2005.08.004
Ihaka, R., Gentleman, R., 1996. R: A Language for Data Analysis and Graphics. J. Comput. Graph. Stat. 5, 299. https://doi.org/10.2307/1390807
Jenkins, S.J., Mountford, A.P., 2005. Dendritic Cells Activated with Products Released by Schistosome Larvae Drive Th2-Type Immune Responses, Which Can Be Inhibited by Manipulation of CD40 Costimulation. Infect. Immun. 73, 395–402. https://doi.org/10.1128/IAI.73.1.395-402.2005
Jensen, K.K., Andreatta, M., Marcatili, P., Buus, S., Greenbaum, J.A., Yan, Z., Sette, A., Peters, B., Nielsen, M., 2018. Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology 154, 394–406. https://doi.org/10.1111/imm.12889
Jespersen, M.C., Peters, B., Nielsen, M., Marcatili, P., 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 45, W24–W29. https://doi.org/10.1093/nar/gkx346
Jiggins, F.M., 2002. Male-Killing Wolbachia and Mitochondrial DNA: Selective Sweeps, Hybrid Introgression and Parasite Population Dynamics 8.

Jonassen, T.M., Imsland, a K., Fitzgerald, R., Bonga, S.W., Ham, E. V, Naevdal, G., Stefánsson, M.O., Stefansson, S.O., 2000. Geographic variation in growth and food conversion efficiency of juvenile Atlantic halibut related to latitude. J. Fish Biol. 56, 279–294. https://doi.org/10.1006/jfbi.1999.1159

Jonsson, B., L'Abee-Lund, J.H., 1993. Latitudinal clines in life-history variables of anadromous brown trout in Europe. J. Fish Biol. 43, 1–16. https://doi.org/10.1111/j.1095-8649.1993.tb01175.x

Junge, C., Vøllestad, L.A., Barson, N.J., Haugen, T.O., Otero, J., Sætre, G.-P., Leder, E.H., Primmer, C.R., 2011. Strong gene flow and lack of stable population structure in the face of rapid adaptation to local temperature in a spring-spawning salmonid, the European grayling (Thymallus thymallus). Heredity 106, 460–471. https://doi.org/10.1038/hdy.2010.160

Karvonen, A., Cheng, G.-H., SeppäLä, O., Valtonen, E.T., 2005. Intestinal distribution and fecundity of two species of Diplostomum parasites in definitive hosts. Parasitology 132, 357. https://doi.org/10.1017/S0031182005009091

Karvonen, A., Paukku, S., Valtonen, E.T., Hudson, P.J., 2003. Transmission, infectivity and survival of Diplostomum spathaceum cercariae. Parasitology 127, 217–224. https://doi.org/10.1017/S0031182003003561

Karvonen, A., Savolainen, M., Seppälä, O., Valtonen, E.T., 2006. Dynamics of Diplostomum spathaceum infection in snail hosts at a fish farm. Parasitol. Res. 99, 341–345. https://doi.org/10.1007/s00436-006-0137-8

Kaufman, J., 2018. Unfinished Business: Evolution of the MHC and the Adaptive Immune System of Jawed Vertebrates 30.

Keeley, E.R., Grant, J.W.A., 2001. Prey size of salmonid fishes in streams, lakes, and oceans. Can. J. Fish. Aquat. Sci. 58, 1122–1132. https://doi.org/10.1139/cjfas-58-6-1122

Keeney, D.B., Waters, J.M., Poulin, R., 2006. Microsatellite loci for the New Zealand trematode Maritrema novaezealandensis: PRIMER NOTE. Mol. Ecol. Notes 6, 1042–1044. https://doi.org/10.1111/j.1471-8286.2006.01426.x

Kelleher, A., Darwiche, R., Rezende, W.C., Farias, L.P., Leite, L.C.C., Schneiter, R., Asojo, O.A., 2014. *Schistosoma mansoni* venom allergen-like protein 4 (SmVAL4) is a novel lipid-binding SCP/TAPS protein that lacks the prototypical CAP motifs. Acta Crystallogr. D Biol. Crystallogr. 70, 2186–2196. https://doi.org/10.1107/S1399004714013315

Kjøglum, S., Larsen, S., Bakke, H.G., Grimholt, U., 2006. How specific MHC class I and class II combinations affect disease resistance against infectious salmon anaemia in Atlantic salmon (Salmo salar). Fish Shellfish Immunol. 21, 431–441. https://doi.org/10.1016/j.fsi.2006.02.001

Klemetsen, A., Amundsen, P.-A., Dempson, J.B., Jonsson, B., Jonsson, N., O'Connell, M.F., Mortensen, E., 2003. Atlantic salmon Salmo salar L., brown trout Salmo trutta L. and Arctic charr Salvelinus alpinus (L.): a review of aspects of their life histories. Ecol. Freshw. Fish 12, 1–59. https://doi.org/10.1034/j.1600-0633.2003.00010.x

Klemetsen, A, P-a, A., Jb, D., Jonsson, B., 2003. Atlantic salmon Salmo salar L ., brown trout Salmo trutta L . and Arctic charr Salvelinus alpinus (L .): a review of aspects of their life histories 1–59.

Knudsen, R., Amundsen, P.-A., Nilsen, R., Kristoffersen, R., Klemetsen, A., 2008. Food borne parasites as indicators of trophic segregation between Arctic charr and brown trout. Environ. Biol. Fishes 83, 107–116. https://doi.org/10.1007/s10641-007-9216-7

Kocher, A., Guilbert, É., Lhuillier, É., Murienne, J., 2015. Sequencing of the mitochondrial genome of the avocado lace bug Pseudacysta perseae (Heteroptera, Tingidae) using a genome skimming approach. C. R. Biol. 338, 149–160. https://doi.org/10.1016/j.crvi.2014.12.004

Kodandaramaiah, U., Simonsen, T.J., Bromilow, S., Wahlberg, N., Sperling, F., 2013. Deceptive single-locus taxonomy and phylogeography: *Wolbachia* -associated divergence in mitochondrial DNA is not reflected in morphology and nuclear markers in a butterfly species. Ecol. Evol. 3, 5167–5176. https://doi.org/10.1002/ece3.886

Kohout, J., 2013. Population genetic structure of brown trout as groundwork for efficient management of fisheries in central European salmonid waters 94.

Koskella, B., Lively, C.M., 2009. Evidence for negative frequency-dependent selection during experimental coevolution of a freshwater snail and a sterilizing trematode. Evolution 63, 2213–2221. https://doi.org/10.1111/j.1558-5646.2009.00711.x

Krause, J., Unger, T., Noçon, A., Malaspinas, A.-S., Kolokotronis, S.-O., Stiller, M., Soibelzon, L., Spriggs, H., Dear, P.H., Briggs, A.W., Bray, S.C., O'Brien, S.J., Rabeder, G., Matheus, P., Cooper, A., Slatkin, M., Pääbo, S., Hofreiter, M., 2008. Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. BMC Evol. Biol. 8, 220. https://doi.org/10.1186/1471-2148-8-220

Kristmundsson, A., Richter, S., 2009. Parasites of resident arctic charr , Salvelinus alpinus , and brown trout , Salmo trutta , in two lakes in Iceland. Icel Agri Sci 22, 5–18.

Krkosek, M., Lewis, M.A., Volpe, J.P., 2005. Transmission dynamics of parasitic sea lice from farm to wild salmon. Proc. R. Soc. B Biol. Sci. 272, 689–696. https://doi.org/10.1098/rspb.2004.3027

Kruger, F.J., Joubert, P.H., 1990. Scanning electron microscopical observations of shedding tegument of schistosoma 3.

Kuchta, R., Brabec, J., Kubáčková, P., Scholz, T., 2013. Tapeworm Diphyllobothrium dendriticum (Cestoda)—Neglected or Emerging Human Parasite? PLoS Negl. Trop. Dis. 7, e2535. https://doi.org/10.1371/journal.pntd.0002535

Kudlai, O., Oros, M., Kostadinova, A., Georgieva, S., 2017a. Exploring the diversity of Diplostomum (Digenea: Diplostomidae) in fishes from the River Danube using mitochondrial DNA barcodes. Parasit. Vectors 10. https://doi.org/10.1186/s13071-017-2518-5

Kudlai, O., Oros, M., Kostadinova, A., Georgieva, S., 2017b. Exploring the diversity of Diplostomum (Digenea: Diplostomidae) in fishes from the River Danube using mitochondrial DNA barcodes. Parasit. Vectors 10. https://doi.org/10.1186/s13071-017-2518-5

Kuhn, J.A., Frainer, A., Knudsen, R., Kristoffersen, R., Amundsen, P.-A., 2016. Effects of fish species composition on *Diphyllobothrium* spp. infections in brown trout -

is three-spined stickleback a key species? J. Fish Dis. 39, 1313–1323. https://doi.org/10.1111/jfd.12467
Kutz, S.J., Jenkins, E.J., Veitch, A.M., Ducrocq, J., Polley, L., Elkin, B., Lair, S., 2009. The Arctic as a model for anticipating, preventing, and mitigating climate change impacts on host–parasite interactions. Vet. Parasitol. 163, 217–228. https://doi.org/10.1016/j.vetpar.2009.06.008
L'Abee-Lund, J.H., Langeland, A., Saegrov, H., 1992. Piscivory by brown trout Salmo trutta L. and Arctic charr Salvelinus alpinus (L.) in Norwegian lakes. J. Fish Biol. 41, 91–101. https://doi.org/10.1111/j.1095-8649.1992.tb03172.x
L'Abée-Lund, J.H., Langeland, A., Saegrov, H., 1992. Piscivory by brown trout *Salmo trutta* L. and Arctic charr *Salvelinus alpinus* (L.) in Norwegian lakes. J. Fish Biol. 41, 91–101. https://doi.org/10.1111/j.1095-8649.1992.tb03172.x
Lamaze, F.C., Pavey, S.A., Normandeau, E., Roy, G., Garant, D., Bernatchez, L., 2014a. Neutral and selective processes shape MHC gene diversity and expression in stocked brook charr populations (Salvelinus fontinalis). Mol. Ecol. 23, 1730–1748. https://doi.org/10.1111/mec.12684
Lamaze, F.C., Pavey, S.A., Normandeau, E., Roy, G., Garant, D., Bernatchez, L., 2014b. Neutral and selective processes shape MHC gene diversity and expression in stocked brook charr populations (Salvelinus fontinalis). Mol. Ecol. 23, 1730–1748. https://doi.org/10.1111/mec.12684
Lamaze, F.C., Pavey, S.A., Normandeau, E., Roy, G., Garant, D., Bernatchez, L., 2014c. Neutral and selective processes shape MHC gene diversity and expression in stocked brook charr populations (*Salvelinus fontinalis*). Mol. Ecol. 23, 1730–1748. https://doi.org/10.1111/mec.12684
Langefors, Åsa, Lohm, J., von Schantz, T., 2001. Allelic polymorphism in MHC class II B in four populations of Atlantic salmon (Salmo salar). Immunogenetics 53, 329–336. https://doi.org/10.1007/s002510100329
Langefors, A., Lohm, J., Grahn, M., Andersen, O., Schantz, T. v., 2001. Association between major histocompatibility complex class IIB alleles and resistance to Aeromonas salmonicida in Atlantic salmon. Proc. R. Soc. B Biol. Sci. 268, 479–485. https://doi.org/10.1098/rspb.2000.1378
Larsen, A.H., Bresciani, J., Buchmann, K., 2005. Pathogenicity of Diplostomum cercariae in rainbow trout, and alternative measures to prevent diplostomosis in fish farms. Bull. Eur. Assoc. Fish Pathol. 25, 20–27.
Lawton, S.P., Bowen, L.I., Emery, A.M., Majoros, G., 2017. Signatures of mito-nuclear discordance in Schistosoma turkestanicum indicate a complex evolutionary history of emergence in Europe. Parasitology 144, 1752–1762. https://doi.org/10.1017/S0031182017000920
Lawton, S.P. and Majoros, G., 2013. A foreign invader or a reclusive native? DNA bar coding reveals a distinct European lineage of the zoonotic parasite Schistosoma turkestanicum (syn. Orientobilharzia turkestanicum ()). *Infection, Genetics and Evolution*, *14*, pp.186-193.
Le, T H, Blair, D., Agatsuma, T., Humair, P.F., Campbell, N.J., Iwagami, M., Littlewood, D.T., Peacock, B., Johnston, D. a, Bartley, J., Rollinson, D., Herniou, E. a, Zarlenga, D.S., McManus, D.P., 2000. Phylogenies inferred from mitochondrial

gene orders-a cautionary tale from the parasitic flatworms. Mol. Biol. Evol. 17, 1123–1125. https://doi.org/10.1093/oxfordjournals.molbev.a026393
Le, T.H., Blair, D., McManus, D.P., 2002. Mitochondrial genomes of parasitic flatworms. Trends Parasitol. 18, 206–213. https://doi.org/10.1016/S1471-4922(02)02252-3
Le, T. H., Blair, D., McManus, D.P., 2000. Mitochondrial DNA sequences of human schistosomes: the current status. Int. J. Parasitol. 30, 283–290. https://doi.org/10.1016/S0020-7519(99)00204-0
Levy, S., Shoham, T., 2005a. Protein-Protein Interactions in the Tetraspanin Web. Physiology 20, 218–224. https://doi.org/10.1152/physiol.00015.2005
Levy, S., Shoham, T., 2005b. The tetraspanin web modulates immune-signalling complexes. Nat. Rev. Immunol. 5, 136–148. https://doi.org/10.1038/nri1548
LEWIS, F.A., WILSON, E.M., 1981. Strain Differences in Lymphocyte Responses and In Vitro Suppressor Cell Induction Between Schistosoma mansoni- infected C57BL/6 and CBA Mice. INFECT IMMUN 32, 8.
Li, L., Yu, L.Y., Zhu, X.Q., Wang, C.R., Zhai, Y.Q., Zhao, J.P., 2008. Orientobilharzia turkestanicum is grouped within African schistosomes based on phylogenetic analyses using sequences of mitochondrial genes. Parasitol. Res. 102, 939–943. https://doi.org/10.1007/s00436-007-0857-4
Littlewood, D.T.J., Lockyer, A.E., Webster, B.L., Johnston, D.A., Le, T.H., 2006. The complete mitochondrial genomes of Schistosoma haematobium and Schistosoma spindale and the evolutionary history of mitochondrial genome changes among parasitic flatworms. Mol. Phylogenet. Evol. 39, 452–467. https://doi.org/10.1016/j.ympev.2005.12.012
Liu, G.-H., Gasser, R.B., Young, N.D., Song, H.-Q., Ai, L., Zhu, X.-Q., 2014. Complete mitochondrial genomes of the 'intermediate form' of Fasciola and Fasciola gigantica, and their comparison with F. hepatica. Parasit. Vectors 7, 150. https://doi.org/10.1186/1756-3305-7-150
Liu, Z.X., Zhang, Y., Liu, Y.T., Chang, Q.C., Su, X., Fu, X., Yue, D.M., Gao, Y., Wang, C.R., 2016. Complete mitochondrial genome of Echinostoma hortense (Digenea: Echinostomatidae). Korean J. Parasitol. 54, 173–179. https://doi.org/10.3347/kjp.2016.54.2.173
Lively, C.M., 1999. Migration, Virulence, and the Geographic Mosaic of Adaptation by Parasites. Am. Nat. 153, S34–S47. https://doi.org/10.1086/303210
Loch, T.P., Faisal, M., 2014. Chryseobacterium aahli sp. nov., isolated from lake trout (Salvelinus namaycush) and brown trout (Salmo trutta), and emended descriptions of chryseobacterium ginsenosidimutans and chryseobacterium gregarium. Int. J. Syst. Evol. Microbiol. 64, 1573–1579. https://doi.org/10.1099/ijs.0.052373-0
Locke, S. a., Daniel McLaughlin, J., Marcogliese, D.J., 2010. DNA barcodes show cryptic diversity and a potential physiological basis for host specificity among Diplostomoidea (Platyhelminthes: Digenea) parasitizing freshwater fishes in the St. Lawrence River, Canada. Mol. Ecol. 19, 2813–2827. https://doi.org/10.1111/j.1365-294X.2010.04713.x
Locke, S.A., Daniel Mclaughlin, J., Marcogliese, D.J., 2010a. DNA barcodes show cryptic diversity and a potential physiological basis for host specificity among Diplostomoidea (Platyhelminthes: Digenea) parasitizing freshwater fishes in the

St. Lawrence River, Canada: Barcoding diplostomoid metacercariae. Mol. Ecol. 19, 2813–2827. https://doi.org/10.1111/j.1365-294X.2010.04713.x
Locke, S.A., Mclaughlin, J.D., Marcogliese, D.J., 2013. Predicting the similarity of parasite communities in freshwater fishes using the phylogeny, Ecology and proximity of hosts. Oikos 122, 73–83. https://doi.org/10.1111/j.1600-0706.2012.20211.x
Locke, S.A., Mclaughlin, J.D., Marcogliese, D.J., 2010b. DNA barcodes show cryptic diversity and a potential physiological basis for host specificity among Diplostomoidea (Platyhelminthes: Digenea) parasitizing freshwater fishes in the St. Lawrence River, Canada. Mol. Ecol. 19, 2813–2827. https://doi.org/10.1111/j.1365-294X.2010.04713.x
Loreille, O., Orlando, L., Patou-Mathis, M., Philippe, M., Taberlet, P., Hänni, C., 2001. Ancient DNA analysis reveals divergence of the cave bear, Ursus spelaeus , and brown bear, Ursus arctos , lineages. Curr. Biol. 11, 200–203. https://doi.org/10.1016/S0960-9822(01)00046-X
Louhi, K.-R., Karvonen, A., Rellstab, C., Jokela, J., 2010. Is the population genetic structure of complex life cycle parasites determined by the geographic range of the most motile host? Infect. Genet. Evol. 10, 1271–1277. https://doi.org/10.1016/j.meegid.2010.08.013
Loukas, A., Tran, M., Pearson, M.S., 2007. Schistosome membrane proteins as vaccines. Int. J. Parasitol. 37, 257–263. https://doi.org/10.1016/j.ijpara.2006.12.001
Lyholt, H., Buchmann, K., 1996. Diplostomum spathaceum:effects of temperature and light on cercarial shedding and infection of rainbow trout. Dis. Aquat. Organ. 25, 169–173. https://doi.org/10.3354/dao025169
Lymbery, a J., Thompson, R.C. a, 2012. The molecular epidemiology of parasite infections: tools and applications. Mol. Biochem. Parasitol. 181, 102–16. https://doi.org/10.1016/j.molbiopara.2011.10.006
Ma, J., He, J.-J., Liu, G.-H., Leontovyč, R., Kašný, M., Zhu, X.-Q., 2016. Complete mitochondrial genome of the giant liver fluke Fascioloides magna (Digenea: Fasciolidae) and its comparison with selected trematodes. Parasit. Vectors 9, 429. https://doi.org/10.1186/s13071-016-1699-7
MacCrimmon, H.R., Marshall, T.L., Gots, B.L., 1970. World Distribution of Brown Trout, *Salmo trutta* : Further Observations. J. Fish. Res. Board Can. 27, 811–818. https://doi.org/10.1139/f70-085
MacKenzie, K., Longshaw, M., Begg, G.S., Mcvicar, A.H., 1998. Sea lice (Copepoda : Caligidae) on wild sea trout (Salmo trutta L .) in Scotland. ICES J. Mar. Sci. 55, 151–162. https://doi.org/10.1006/jmsc.1997.0362
Maizels, R.M., 2009. Parasite immunomodulation and polymorphisms of the immune system. J. Biol. 8, 62. https://doi.org/10.1186/jbiol166
Malé, P.-J.G., Bardon, L., Besnard, G., Coissac, E., Delsuc, F., Engel, J., Lhuillier, E., Scotti-Saintagne, C., Tinaut, A., Chave, J., 2014. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Mol. Ecol. Resour. n/a-n/a. https://doi.org/10.1111/1755-0998.12246
Malmstrøm, M., Matschiner, M., Tørresen, O.K., Star, B., Snipen, L.G., Hansen, T.F., Baalsrud, H.T., Nederbragt, A.J., Hanel, R., Salzburger, W., Stenseth, N.C., Jakobsen, K.S., Jentoft, S., 2016. Evolution of the immune system influences

speciation rates in teleost fishes. Nat. Genet. 48, 1204–1210. https://doi.org/10.1038/ng.3645
Manier, M.K., Arnold, S.J., 2005. Population genetic analysis identifies source-sink dynamics for two sympatric garter snake species (Thamnophis elegans and Thamnophis sirtalis): source-sink dynamics for two garter snakes. Mol. Ecol. 14, 3965–3976. https://doi.org/10.1111/j.1365-294X.2005.02734.x
Marcogliese, D.J., 2008. The impact of climate change on the parasites and infectious diseases of aquatic animals. Rev. Sci. Tech. Int. Off. Epizoot. 27, 467–484.
Markova, A.K., Puzachenko, A.Y., van Kolfschoten, T., 2010. The North Eurasian mammal assemblages during the end of MIS 3 (Brianskian–Late Karginian–Denekamp Interstadial). Quat. Int. 212, 149–158. https://doi.org/10.1016/j.quaint.2009.02.010
Matthews, B., Harmon, L.J., M'Gonigle, L., Marchinko, K.B., Schaschl, H., 2010. Sympatric and Allopatric Divergence of MHC Genes in Threespine Stickleback. PLoS ONE 5, e10948. https://doi.org/10.1371/journal.pone.0010948
McKeown, N.J., Hynes, R.A., Duguid, R.A., Ferguson, A., Prodöhl, P.A., 2010. Phylogeographic structure of brown trout Salmo trutta in Britain and Ireland: Glacial refugia, postglacial colonization and origins of sympatric populations. J. Fish Biol. 76, 319–347. https://doi.org/10.1111/j.1095-8649.2009.02490.x
McManus, D.P., Loukas, A., 2008. Current Status of Vaccines for Schistosomiasis. Clin. Microbiol. Rev. 21, 225–242. https://doi.org/10.1128/CMR.00046-07
McSorley, H.J., Hewitson, J.P., Maizels, R.M., 2013. Immunomodulation by helminth parasites: Defining mechanisms and mediators. Int. J. Parasitol. 43, 301–310. https://doi.org/10.1016/j.ijpara.2012.11.011
Medel, R., Mendez, M. a., Ossa, C.G., Botto-Mahan, C., 2010. Arms Race Coevolution: The Local and Geographical Structure of a Host–Parasite Interaction. Evol. Educ. Outreach 3, 26–31. https://doi.org/10.1007/s12052-009-0191-7
Mercado, R., Yamasaki, H., Kato, M., Muñoz, V., Sagua, H., Torres, P., Castillo, D., 2010. Molecular identification of the Diphyllobothrium species causing diphyllobothriasis in Chilean patients. Parasitol. Res. 106, 995–1000. https://doi.org/10.1007/s00436-010-1765-6
Micallef, G., Bickerdike, R., Reiff, C., Fernandes, J.M.O., Bowman, A.S., Martin, S.A.M., 2012. Exploring the Transcriptome of Atlantic Salmon (Salmo salar) Skin, a Major Defense Organ. Mar. Biotechnol. 14, 559–569. https://doi.org/10.1007/s10126-012-9447-2
Miller, K.M., Withler, R.E., 2004. Mhc Diversity in Pacific Salmon: Population Structure and Trans-Species Allelism. Hereditas 127, 83–95. https://doi.org/10.1111/j.1601-5223.1997.00083.x
Milot, E., Perrier, C., Papillon, L., Dodson, J.J., Bernatchez, L., 2013. Reduced fitness of Atlantic salmon released in the wild after one generation of captive breeding. Evol. Appl. 6, 472–485. https://doi.org/10.1111/eva.12028
Mittelbrunn, M., Yanez-Mo, M., Sancho, D., Sanchez-Madrid, F., n.d. Cutting Edge: Dynamic Redistribution of Tetraspanin CD81 at the Central Zone of the Immune Synapse in Both T Lymphocytes and APC. J. Immunol. 5.
Mjaaland, S., Markussen, T., Sindre, H., Kjøglum, S., Dannevig, B.H., Larsen, S., Grimholt, U., 2005. Susceptibility and immune responses following experimental

infection of MHC compatible Atlantic salmon (Salmo salar L.) with different infectious salmon anaemia virus isolates. Arch. Virol. 150, 2195–2216. https://doi.org/10.1007/s00705-005-0588-8
Molloy, A.S., Holland, C., Poole, R., Biology, S., Proceedings, E., Irish, R., Poole, R., Molloy, S., Holland, C., 2014. Parasites of brown and salmo trutta l . from the coast 93, 137–142.
Molloy, S., Holland, C., Poole, R., 1995. Metazoan parasite community structure in brown trout from two lakes in western Ireland. J. Helminthol. 69, 237. https://doi.org/10.1017/S0022149X0001419X
Moné, H., Holtfreter, M.C., Allienne, J.-F., Mintsa-Nguéma, R., Ibikounlé, M., Boissier, J., Berry, A., Mitta, G., Richter, J., Mouahid, G., 2015. Introgressive hybridizations of Schistosoma haematobium by Schistosoma bovis at the origin of the first case report of schistosomiasis in Corsica (France, Europe). Parasitol. Res. 114, 4127–4133. https://doi.org/10.1007/s00436-015-4643-4
Monzón-Argüello, C., Garcia de Leaniz, C., Gajardo, G., Consuegra, S., 2014. Eco-immunology of fish invasions: the role of MHC variation. Immunogenetics 66, 393–402. https://doi.org/10.1007/s00251-014-0771-8
Monzón-Argüello, C, Garcia de Leaniz, C., Gajardo, G., Consuegra, S., 2014. Eco-immunology of fish invasions: the role of MHC variation. Immunogenetics 66, 393–402. https://doi.org/10.1007/s00251-014-0771-8
Moody, J., Gaten, E., 1982. The population dynamics of eyeflukes Diplostomum spathaceum and Tylodelphys clavata (Digenea: Diplostomatidae) in rainbow and brown trout in Rutland Water: 1974-1978. Hydrobiologia 88, 207–209. https://doi.org/10.1007/BF00008315
Morales, H.E., Pavlova, A., Joseph, L., Sunnucks, P., 2015. Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. Mol. Ecol. 24, 2820–2837. https://doi.org/10.1111/mec.13203
Morand, S., Guegan, J.-F., 2000. Distribution and abundance of parasite nematodes: ecological specialisation, phylogenetic constraint or simply epidemiology? Oikos 88, 563–573. https://doi.org/10.1034/j.1600-0706.2000.880313.x
Morgan, A.D., Gandon, S., Buckling, A., 2005. The effect of migration on local adaptation in a coevolving host–parasite system. Nature 437, 253–256. https://doi.org/10.1038/nature03913
Moszczynska, A., Locke, S.A., McLaughlin, J.D., Marcogliese, D.J., Crease, T.J., 2009a. Development of primers for the mitochondrial cytochrome c oxidase i gene in digenetic trematodes (Platyhelminthes) illustrates he challenge of barcoding parasitic helminths. Mol. Ecol. Resour. 9, 75–82. https://doi.org/10.1111/j.1755-0998.2009.02634.x
Moszczynska, A., Locke, S.A., McLaughlin, J.D., Marcogliese, D.J., Crease, T.J., 2009b. Development of primers for the mitochondrial cytochrome *c* oxidase I gene in digenetic trematodes (Platyhelminthes) illustrates the challenge of barcoding parasitic helminths. Mol. Ecol. Resour. 9, 75–82. https://doi.org/10.1111/j.1755-0998.2009.02634.x
Mudry, D.R., Stewart Anderson, R., 1977. Helminth and arthropod parasites of fishes in the mountain national parks of Canada. J. Fish Biol. 11, 21–33. https://doi.org/10.1111/j.1095-8649.1977.tb04095.x

Nabi, A.S., Tanveer, S., Ganie, S.A., Ahad, S., Niyaz, U., 2016. Acanthocephalan infestation in fishes – A review 2, 33–38.
Nachev, M., Sures, B., 2016. Environmental parasitology: Parasites as accumulation bioindicators in the marine environment. J. Sea Res. 113, 45–50. https://doi.org/10.1016/j.seares.2015.06.005
Natsopoulou, M.E., 2010. Parasites and parallel divergence of individual MHC allelic richness in Icelandic threespine stickleback (Gasterosteus aculeatus L.): Contrasting habitats and population divergence.
Ndeda, V.M., Owiti, D.O., Aketch, B.O., Onyango, D.M., 2013. Genetic Relatedness of Diplostomum Species (Digenea: Diplostomidae) Infesting Nile Tilapia (Oreochromis Niloticus L.) in Western Kenya. Open J. Appl. Sci. 03, 441–448. https://doi.org/10.4236/ojapps.2013.38055
Neil, P., 2013. Life history strategies of brown trout (salmo trutta) current knowledge and research gaps a review of the scientific literature Travis van Leeuwen (University of Glasgow) Supervisors Prof . Colin Adams (University of Glasgow) Dr . Patrick Boylan (L.
Nelson, P.A., Dick, T.A., 2002. Factors shaping the parasite communities of trout-perch, *Percopsis omiscomaycus* Walbaum (Osteichthyes: Percopsidae), and the importance of scale. Can. J. Zool. 80, 1986–1999. https://doi.org/10.1139/z02-188
N'Goran, E., Brémond, P., Sellin, E., Sellin, B., Théron, A., 1997. Intraspecific diversity of Schistosoma haematobium in West Africa: chronobiology of cercarial emergence. Acta Trop. 66, 35–44. https://doi.org/10.1016/S0001-706X(97)00676-1
Nilsen, E.B., Milner-Gulland, E.., Schofield, L., Mysterud, A., Stenseth, N.C., Coulson, T., 2007. Wolf reintroduction to Scotland: public attitudes and consequences for red deer management. Proc. R. Soc. B Biol. Sci. 274, 995–1003. https://doi.org/10.1098/rspb.2006.0369
Nordling, D., 1998. Reproductive effort reduces speci®c immune response and parasite resistance 8.
O'Brien, S.H., Wilson, L.J., Webb, A., Cranswick, P.A., 2008. Revised estimate of numbers of wintering red-throated divers Gavia stellata in Great Britain. Bird Study 55, 152–160. https://doi.org/10.1080/00063650809461517
O'Farrell, B., Benzie, J. a H., McGinnity, P., de Eyto, E., Dillane, E., Coughlan, J., Cross, T.F., 2013a. Selection and phylogenetics of salmonid MHC class I: wild brown trout (Salmo trutta) differ from a non-native introduced strain. PloS One 8, e63035. https://doi.org/10.1371/journal.pone.0063035
O'Farrell, B., Benzie, J.A.H., McGinnity, P., de Eyto, E., Dillane, E., Coughlan, J., Cross, T.F., 2013b. Selection and Phylogenetics of Salmonid MHC Class I: Wild Brown Trout (Salmo trutta) Differ from a Non-Native Introduced Strain. PLoS ONE 8, e63035. https://doi.org/10.1371/journal.pone.0063035
Olson, P.D., Cribb, T.H., Tkach, V. V., Bray, R.A., Littlewood, D.T.J., 2003. Phylogeny and classification of the Digenea (Platyhelminthes: Trematoda). Int. J. Parasitol. 33, 733–755. https://doi.org/10.1016/S0020-7519(03)00049-3
Oscoz, J., Leunda, P.M., Campos, F., Escala, M.C., Miranda, R., n.d. Diet of 0+ brown trout (Salmo trutta L., 1758) from the river Erro (Navarra, north of Spain) 8.

Palstra, F.P., O'Connell, M.F., Ruzzante, D.E., 2007. Population structure and gene flow reversals in Atlantic salmon (*Salmo salar*) over contemporary and long-term temporal scales: effects of population size and life history: GENE FLOW IN ATLANTIC SALMON. Mol. Ecol. 16, 4504–4522. https://doi.org/10.1111/j.1365-294X.2007.03541.x

Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20, 289–290. https://doi.org/10.1093/bioinformatics/btg412

Park, G.-M., 2007. Genetic comparison of liver flukes, Clonorchis sinensis and Opisthorchis viverrini, based on rDNA and mtDNA gene sequences. Parasitol. Res. 100, 351–357. https://doi.org/10.1007/s00436-006-0269-x

Pastor-Valle, J., González, L.M., Martín-Clemente, J.P., Merino, F.J., Gottstein, B., Gárate, T., 2014. Molecular diagnosis of diphyllobothriasis in Spain, most presumably acquired via imported fish, or sojourn abroad. New Microbes New Infect. 2, 1–6. https://doi.org/10.1002/2052-2975.28

Patten, M.M., Carioscia, S.A., Linnen, C.R., 2015. Biased introgression of mitochondrial and nuclear genes: a comparison of diploid and haplodiploid systems. Mol. Ecol. 24, 5200–5210. https://doi.org/10.1111/mec.13318

Payseur, B.A., Rieseberg, L.H., 2016. A genomic perspective on hybridization and speciation. Mol. Ecol. 25, 2337–2360. https://doi.org/10.1111/mec.13557

Peduzzi, R., Boucher-Rodoni, R., 2001. Resurgence of human bothriocephalosis (Diphyllobothrium latum) in the subalpine lake region. J. Limnol. 60. https://doi.org/10.4081/jlimnol.2001.41

Penn, D.J., 2002. The Scent of Genetic Compatibility: Sexual Selection and the Major Histocompatibility Complex. Ethology 108, 1–21. https://doi.org/10.1046/j.1439-0310.2002.00768.x

Penn, D.J., Damjanovich, K., Potts, W.K., 2002. MHC heterozygosity confers a selective advantage against multiple-strain infections. Proc. Natl. Acad. Sci. 99, 11260–11264. https://doi.org/10.1073/pnas.162006499

Perea, S., Vukić, J., Šanda, R., Doadrio, I., 2016. Ancient Mitochondrial Capture as Factor Promoting Mitonuclear Discordance in Freshwater Fishes: A Case Study in the Genus Squalius (Actinopterygii, Cyprinidae) in Greece. PLOS ONE 11, e0166292. https://doi.org/10.1371/journal.pone.0166292

Pérez-del-Olmo, A., Georgieva, S., Pula, H.J., Kostadinova, A., 2014a. Molecular and morphological evidence for three species of Diplostomum (Digenea: Diplostomidae), parasites of fishes and fish-eating birds in Spain 15.

Pérez-del-Olmo, A., Georgieva, S., Pula, H.J., Kostadinova, A., 2014b. Molecular and morphological evidence for three species of Diplostomum (Digenea: Diplostomidae), parasites of fishes and fish-eating birds in Spain. Parasit. Vectors 7, 502. https://doi.org/10.1186/s13071-014-0502-x

Pérez-Sánchez, R., Valero, M.L., Ramajo-Hernández, A., Siles-Lucas, M., Ramajo-Martín, V., Oleaga, A., 2008. A proteomic approach to the identification of tegumental proteins of male and female Schistosoma bovis worms. Mol. Biochem. Parasitol. 161, 112–123. https://doi.org/10.1016/j.molbiopara.2008.06.011

Peters, M.B., Turner, T.F., 2008. Genetic variation of the major histocompatibility complex (MHC class II β gene) in the threatened Gila trout, Oncorhynchus gilae gilae. Conserv. Genet. 9, 257–270. https://doi.org/10.1007/s10592-007-9336-9
Petre, S.J., Bonar, S.A., 2017. Determination of Habitat Requirements for Apache Trout. Trans. Am. Fish. Soc. 146, 1–15. https://doi.org/10.1080/00028487.2016.1225597
Piasecki, W., Goodwin, A.E., Eiras, J.C., Nowak, B.F., 2004. Importance of Copepoda in Freshwater Aquaculture. Zool. Stud. 13.
Piertney, S.B., Oliver, M.K., 2006. The evolutionary ecology of the major histocompatibility complex. Heredity 96, 7–21. https://doi.org/10.1038/sj.hdy.6800724
Pinto, H.A., Mati, V.L.T., Melo, A.L., 2016. Can the same species of Platynosomum (Trematoda: Dicrocoeliidae) infect both mammalian and avian hosts? J. Helminthol. 90, 372–376. https://doi.org/10.1017/S0022149X15000152
Pinto, H.A., Pulido-Murillo, E.A., Braga, R.R., Mati, V.L.T., Melo, A.L., Tkach, V.V., 2018. DNA sequences confirm low specificity to definitive host and wide distribution of the cat pathogen Platynosomum illiciens (= P. fastosum) (Trematoda: Dicrocoeliidae). Parasitol. Res. 117, 1975–1978. https://doi.org/10.1007/s00436-018-5866-y
Pleass, R.J., Kusel, J.R., Woof, J.M., 2000. Cleavage of Human IgE Mediated by *Schistosoma mansoni*. Int. Arch. Allergy Immunol. 121, 194–204. https://doi.org/10.1159/000024317
Poulin, R., 2016. Greater diversification of freshwater than marine parasites of fish. Int. J. Parasitol. 46, 275–279. https://doi.org/10.1016/j.ijpara.2015.12.002
Poulin, R., 2007. Are there general laws in parasite ecology? Parasitology 134, 763. https://doi.org/10.1017/S0031182006002150
Poulin, R., Mouillot, D., 2003. Host introductions and the geography of parasite taxonomic diversity. J. Biogeogr. 30, 837–845. https://doi.org/10.1046/j.1365-2699.2003.00868.x
Preidis, G.A., Hotez, P.J., 2015. The Newest "Omics"—Metagenomics and Metabolomics—Enter the Battle against the Neglected Tropical Diseases. PLoS Negl. Trop. Dis. 9, e0003382. https://doi.org/10.1371/journal.pntd.0003382
Primmer, C.R., 2011. Genetics of local adaptation in salmonid fishes. Heredity 106, 401–403. https://doi.org/10.1038/hdy.2010.158
Prugnolle, F., Liu, H., de Meeûs, T., Balloux, F., 2005. Population genetics of complex life-cycle parasites: an illustration with trematodes. Int. J. Parasitol. 35, 255–263. https://doi.org/10.1016/j.ijpara.2004.10.027
Radwan, J., Zagalska-Neubauer, M., Cichoń, M., Sendecka, J., Kulma, K., Gustafsson, L., Babik, W., 2012. MHC diversity, malaria and lifetime reproductive success in collared flycatchers: MHC and fitness in flycatchers. Mol. Ecol. 21, 2469–2479. https://doi.org/10.1111/j.1365-294X.2012.05547.x
Rahkonen, R, Aalto, J., Koski, P., Särkkä, J., Juntunen, K., 1996. Cestode larvae Diphyllobothrium dendriticum as a cause of heart disease leading to mortality in hatchery-reared sea trout and brown trout. Dis. Aquat. Organ. 25, 15–22. https://doi.org/10.3354/dao025015
Rahkonen, Riitta, Aalto, J., Koski, P., Särkkä, J., Juntunen, K., 1996. Cestode larvae Diphyllobothrium dendriticum as a cause of heart disease leading to mortality in

hatchery-reared sea trout and brown trout. Dis. Aquat. Organ. 25, 15–22. https://doi.org/10.3354/dao025015
Robinson, M.W., Menon, R., Donnelly, S.M., Dalton, J.P., Ranganathan, S., 2009. An Integrated Transcriptomics and Proteomics Analysis of the Secretome of the Helminth Pathogen *Fasciola hepatica*: Proteins associated with invasion and infection of the mammalian host. Mol. Cell. Proteomics 8, 1891–1907. https://doi.org/10.1074/mcp.M900045-MCP200
Rosenkranz, M., Poulin, R., Selbach, C., 2018. Behavioural impacts of trematodes on their snail host: Species-specific effects or generalised response? Ethology 124, 790–795. https://doi.org/10.1111/eth.12808
Rozas, Marco, et al. "First molecular identification of Diphyllobothrium dendriticum plerocercoids from feral rainbow trout (Oncorhynchus mykiss) in Chile." *The Journal of parasitology* 98.6 (2012): 1220-1226.
Ruzycki, J.R., Beauchamp, D.A., Yule, D.L., 2003. Effects of introduced lake trout on native cutthroat trout in yellowstone lake. Ecol. Appl. 13, 23–37. https://doi.org/10.1890/1051-0761(2003)013[0023:EOILTO]2.0.CO;2
Santos, F., Faro, L. de, 2005. The first confirmed case of Diphyllobothrium latum in Brazil. Mem. Inst. Oswaldo Cruz 100, 585–586. https://doi.org/10.1590/S0074-02762005000600013
Saarma, U., Ho, S.Y., Pybus, O.G., Kaljuste, M., Tumanov, I.L., Kojola, I., Vorobiev, A.A., Markov, N.I., Saveljev, A.P., Valdmann, H. and Lyapunova, E.A., 2007. Mitogenetic structure of brown bears (Ursus arctos L.) in northeastern Europe and a new time frame for the formation of European brown bear lineages. *Molecular Ecology*, *16*(2), pp.401-413.
Sato, A., 2003. Origin and Speciation of Haplochromine Fishes in East African Crater Lakes Investigated by the Analysis of Their mtDNA, Mhc Genes, and SINEs. Mol. Biol. Evol. 20, 1448–1462. https://doi.org/10.1093/molbev/msg151
Schad, J., Ganzhorn, J.U., Sommer, S., 2005. Parasite burden and constitution of major histocompatibility complex in the malagasy mouse lemur, microcebus murinus. Evolution 59, 439–450. https://doi.org/10.1111/j.0014-3820.2005.tb01002.x
Schenekar, T., Weiss, S., 2017. Selection and genetic drift in captive versus wild populations: an assessment of neutral and adaptive (MHC-linked) genetic variation in wild and hatchery brown trout (Salmo trutta) populations. Conserv. Genet. 0, 1–12. https://doi.org/10.1007/s10592-017-0949-3
Schluter, D., 1996. Ecological speciation in postglacial fishes 8.
Scholz, T., Garcia, H.H., Kuchta, R., Wicht, B., 2009. Update on the Human Broad Tapeworm (Genus Diphyllobothrium), Including Clinical Relevance. Clin. Microbiol. Rev. 22, 146–160. https://doi.org/10.1128/CMR.00033-08
Scholz, T., Kuchta, R., 2016. Fish-borne, zoonotic cestodes (Diphyllobothrium and relatives) in cold climates: A never-ending story of neglected and (re)-emergent parasites. Food Waterborne Parasitol. 4, 23–38. https://doi.org/10.1016/j.fawpar.2016.07.002
Seifertová, M., Jarkovský, J., Šimková, A., 2016. Does the parasite-mediated selection drive the MHC class IIB diversity in wild populations of European chub (Squalius cephalus)? Parasitol. Res. 115, 1401–1415. https://doi.org/10.1007/s00436-015-4874-4

Selbach, C., Soldánová, M., Georgieva, S., Kostadinova, A., Sures, B., 2015a. Integrative taxonomic approach to the cryptic diversity of Diplostomum spp. in lymnaeid snails from Europe with a focus on the 'Diplostomum mergi' species complex. Parasit. Vectors 8. https://doi.org/10.1186/s13071-015-0904-4
Selbach, C., Soldánová, M., Georgieva, S., Kostadinova, A., Sures, B., 2015b. Integrative taxonomic approach to the cryptic diversity of Diplostomum spp. in lymnaeid snails from Europe with a focus on the 'Diplostomum mergi' species complex. Parasit. Vectors 8. https://doi.org/10.1186/s13071-015-0904-4
Sheath, D.J., Williams, C.F., Reading, A.J., Robert Britton, J., 2015. Parasites of non-native freshwater fishes introduced into england and wales suggest enemy release and parasite acquisition. Biol. Invasions 17, 2235–2246. https://doi.org/10.1007/s10530-015-0857-8
Shekhovtsov, S.V., Katokhin, A.V., Kolchanov, N.A., Mordvinov, V.A., 2010. The complete mitochondrial genomes of the liver flukes Opisthorchis felineus and Clonorchis sinensis (Trematoda). Parasitol. Int. 59, 100–103. https://doi.org/10.1016/j.parint.2009.10.012
Simpson, A.J.G., Schryer, M.D., Cesari, I.M., Evans, W.H., Smithers, S.R., 1981. Isolation and partial characterization of the tegumental outer membrane of adult Schistosoma mansoni. Parasitology 83, 163. https://doi.org/10.1017/S0031182000050137
Simpson, E., 1988. Function of the MHC. Immunol. Suppl. 1, 27–30.
Siwertsson, A., Refsnes, B., Frainer, A., Amundsen, P.A., Knudsen, R., 2016. Divergence and parallelism of parasite infections in Arctic charr morphs from deep and shallow lake habitats. Hydrobiologia 783, 131–143. https://doi.org/10.1007/s10750-015-2563-z
Skelly, P.J., Alan Wilson, R., 2006. Making Sense of the Schistosome Surface, in: Advances in Parasitology. Elsevier, pp. 185–284. https://doi.org/10.1016/S0065-308X(06)63003-0
Smith, N.F., 2001. Spatial heterogeneity in recruitment of larval trematodes to snail intermediate hosts. Oecologia 127, 115–122. https://doi.org/10.1007/s004420000560
Smouse, P.E., Long, J.C., Sokal, R.R., 1986. Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. Syst. Zool. 35, 627. https://doi.org/10.2307/2413122
Sommer, S., 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front. Zool. 2, 16. https://doi.org/10.1186/1742-9994-2-16
Sorrentino, R.P., Carton, Y., Govind, S., 2002. Cellular Immune Response to Parasite Infection in the Drosophila Lymph Gland Is Developmentally Regulated. Dev. Biol. 243, 65–80. https://doi.org/10.1006/dbio.2001.0542
Stables, J.N., Chappell, L.H., 1986a. Diplostomum spathaceum (Rud. 1819): effects of physical factors on the infection of rainbow trout (Salmo gairdneri) by cercariae. Parasitology 93, 71. https://doi.org/10.1017/S0031182000049830
Stables, J.N., Chappell, L.H., 1986b. The epidemiology of diplostomiasis in farmed rainbow trout from north-east Scotland. Parasitology 92, 699. https://doi.org/10.1017/S0031182000065550

Stiller, M., Molak, M., Prost, S., Rabeder, G., Baryshnikov, G., Rosendahl, W., Münzel, S., Bocherens, H., Grandal-d'Anglade, A., Hilpert, B. and Germonpré, M., 2014. Mitochondrial DNA diversity and evolution of the Pleistocene cave bear complex. *Quaternary International*, *339*, pp.224-231.

Stumbo, A.D., Poulin, R., 2016. Possible mechanism of host manipulation resulting from a diel behaviour pattern of eye-dwelling parasites? Parasitology 143, 1261–1267. https://doi.org/10.1017/S0031182016000810

Szidat, L., 1969. Structure, Development, and Behaviour of New Strigeatoid Metacercariae from Subtropical Fishes of South America. J. Fish. Res. Board Can. 26, 753–786. https://doi.org/10.1139/f69-074

Taberlet, P., Bouvet, J., 1994. Mitochondrial DNA Polymorphism, Phylogeography, and Conservation Genetics of the Brown Bear Ursus arctos in Europe. Proc. R. Soc. B Biol. Sci. 255, 195–200. https://doi.org/10.1098/rspb.1994.0028

Tarleton, R.L., Kissinger, J., 2014. Parasite genomics: current status and future prospects 8.

Tarrant, J.M., Robb, L., van Spriel, A.B., Wright, M.D., 2003. Tetraspanins: molecular organisers of the leukocyte surface. Trends Immunol. 24, 610–617. https://doi.org/10.1016/j.it.2003.09.011

Taylor, E.B., 1991. A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. Aquaculture 98, 185–207. https://doi.org/10.1016/0044-8486(91)90383-i

Telford, M.J., Herniou, E.A., Russell, R.B., Littlewood, D.T.J., 2000. Changes in mitochondrial genetic codes as phylogenetic characters: Two examples from the flatworms. Proc. Natl. Acad. Sci. 97, 11359–11364. https://doi.org/10.1073/pnas.97.21.11359

Teitelbaum, C.S., Huang, S., Hall, R.J. and Altizer, S., 2018. Migratory behaviour predicts greater parasite diversity in ungulates. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1875), p.20180089.

Thanchomnang, T., Tantrawatpan, C., Intapan, P.M., Sanpool, O., Lulitanond, V., Tourtip, S., Yamasaki, H., Maleewong, W., 2016. Rapid identification of nine species of diphyllobothriidean tapeworms by pyrosequencing. Sci. Rep. 6. https://doi.org/10.1038/srep37228

Thieltges, D.W., Reise, K., Prinz, K., Jensen, K.T., 2009. Invaders interfere with native parasite-host interactions. Biol. Invasions 11, 1421–1429. https://doi.org/10.1007/s10530-008-9350-y

Tonteri, A., VasemÄgi, A., Lumme, J., Primmer, C.R., 2010. Beyond MHC: Signals of elevated selection pressure on Atlantic salmon (Salmo salar) immune-relevant loci. Mol. Ecol. 19, 1273–1282. https://doi.org/10.1111/j.1365-294X.2010.04573.x

Torres, P., Cuevas, C., Tang, M., Barra, M., Franjola, R., Navarrete, N., Montefusco, A., Otth, L., Wilson, G., Puga, S. and Figueroa, L., 2004. Introduced and native fishes as infection foci of Diphyllobothrium spp. in humans and dogs from two localities at Lake Panguipulli in Southern Chile. *Comparative parasitology*, *71*(2), pp.111-118.

Tran, M.H., Pearson, M.S., Bethony, J.M., Smyth, D.J., Jones, M.K., Duke, M., Don, T.A., McManus, D.P., Correa-Oliveira, R., Loukas, A., 2006. Tetraspanins

on the surface of Schistosoma mansoni are protective antigens against schistosomiasis. Nat. Med. 12, 835–840. https://doi.org/10.1038/nm1430
Trieu, N., Cutmore, S.C., Miller, T.L., Cribb, T.H., 2015. A species pair of Bivesicula Yamaguti, 1934 (Trematoda: Bivesiculidae) in unrelated Great Barrier Reef fishes: implications for the basis of speciation in coral reef fish trematodes. Syst. Parasitol. 91, 231–239. https://doi.org/10.1007/s11230-015-9576-x
Valdiosera, C.E., GarcíA, N., Anderung, C., DaléN, L., CréGut-Bonnoure, E., Kahlke, R.-D., Stiller, M., BrandströM, M., Thomas, M.G., Arsuaga, J.L., GöTherströM, A., Barnes, I., 2007. Staying out in the cold: glacial refugia and mitochondrial DNA phylogeography in ancient European brown bears. Mol. Ecol. 16, 5140–5148. https://doi.org/10.1111/j.1365-294X.2007.03590.x
Valtonen, E.T., Pulkkinen, K., Poulin, R., Julkunen, M., 2001. The structure of parasite component communities in brackish water fishes of the northeastern Baltic Sea. Parasitology 122. https://doi.org/10.1017/S0031182001007491
Van Hellemond, J.J., Retra, K., Brouwers, J.F.H.M., van Balkom, B.W.M., Yazdanbakhsh, M., Shoemaker, C.B., Tielens, A.G.M., 2006. Functions of the tegument of schistosomes: Clues from the proteome and lipidome. Int. J. Parasitol. 36, 691–699. https://doi.org/10.1016/j.ijpara.2006.01.007
van Paridon, B.J., Colwell, D.D., Goater, C.P., Gilleard, J.S., 2017. Population genetic analysis informs the invasion history of the emerging trematode Dicrocoelium dendriticum into Canada. Int. J. Parasitol. 47, 845–856. https://doi.org/10.1016/j.ijpara.2017.04.006
Vehanen, T., Maki-Petays, A., Aspi, J., Muotka, T., 1999. Intercohort competition causes spatial segregation in brown trout in artificial streams. J. Fish Biol. 55, 35–46. https://doi.org/10.1111/j.1095-8649.1999.tb00654.x
Voutilainen, A., van Ooik, T., Puurtinen, M., Kortet, R., Taskinen, J., 2009. Relationship between prevalence of trematode parasite Diplostomum sp. and population density of its snail host Lymnaea stagnalis in lakes and ponds in Finland. Aquat. Ecol. 43, 351–357. https://doi.org/10.1007/s10452-008-9203-x
Waples, R.S., Hendry, A.P., 2008. Special Issue: Evolutionary perspectives on salmonid conservation and management. Evol. Appl. 1, 183–8. https://doi.org/10.1111/j.1752-4571.2008.00035.x
Webster, B.L., Diaw, O.T., Seye, M.M., Webster, J.P., Rollinson, D., 2013. Introgressive Hybridization of Schistosoma haematobium Group Species in Senegal: Species Barrier Break Down between Ruminant and Human Schistosomes. PLoS Negl. Trop. Dis. 7, e2110. https://doi.org/10.1371/journal.pntd.0002110
Whitmarsh, D., Wattage, P., 2006. Public attitudes towards the environmental impact of salmon aquaculture in Scotland. Eur. Environ. 16, 108–121. https://doi.org/10.1002/eet.406
Whyte, S.K., Allan, J.C., Secombes, C.J., Chappell, L.H., 1987. Cercariae and diplostomules of Diplostomum spathaceum (Digenea) elicit an immune response in rainbow trout, Salmo gairdneri Richardson. J. Fish Biol. 31, 185–190. https://doi.org/10.1111/j.1095-8649.1987.tb05311.x
Whyte, S.K., Chappell, L.H., Secombes, C.J., 1990. Protection of rainbow trout, Oncorhynchus mykiss (Richardson), against Diplostomum spathaceum (Digenea):

the role of specific antibody and activated macrophages. J. Fish Dis. 13, 281–291. https://doi.org/10.1111/j.1365-2761.1990.tb00784.x
Whyte, S.K., Secombes, C.J., Chappell, L.H., 1991. Studies on the infectivity of Diplostomum spathaceum in rainbow trout (Oncorhynchus mykiss). J. Helminthol. 65, 169. https://doi.org/10.1017/S0022149X0001066X
Wicht, B., de Marval, F., Gottstein, B., Peduzzi, R., 2008. Imported diphyllobothriasis in Switzerland: molecular evidence of Diphyllobothrium dendriticum (Nitsch, 1824). Parasitol. Res. 102, 201–204. https://doi.org/10.1007/s00436-007-0745-y
Wicht, B., Yanagida, T., Scholz, T., Ito, A., Jimenez, J.A., Brabec, J., 2010. Multiplex PCR for Differential Identification of Broad Tapeworms (Cestoda: Diphyllobothrium) Infecting Humans. J. Clin. Microbiol. 48, 3111–3116. https://doi.org/10.1128/JCM.00445-10
Wieczorek, M., Abualrous, E.T., Sticht, J., Álvaro-Benito, M., Stolzenberg, S., Noé, F., Freund, C., 2017. Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. Front. Immunol. 8. https://doi.org/10.3389/fimmu.2017.00292
Wiemers, M., Fiedler, K., 2007. Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). Front. Zool. 4, 8. https://doi.org/10.1186/1742-9994-4-8
Wilbur, H.M., 1980. Complex Life Cycles. Annu. Rev. Ecol. Syst. 11, 67–93. https://doi.org/10.1146/annurev.es.11.110180.000435
WOOTTEN, R., SMITH, J.W., 1975. Observational and experimental studies on the acquisition of a~~sak~s sp. larvae (nematoda: ascaridida) by trout in fresh water 6.
Xiong, F., Li, W.X., Wu, S.G., Zou, H., Wang, G.T., 2013. Molecular phylogeny and host specificity of the larval Eustrongylides (Nematoda: Dioctophmidae) from freshwater fish in China. J. Parasitol. 99, 137–44. https://doi.org/10.1645/GE-3163.1
Yamasaki, H., Kuramochi, T., 2009. A case of Diphyllobothrium nihonkaiense infection possibly linked to salmon consumption in New Zealand. Parasitol. Res. 105, 583–586. https://doi.org/10.1007/s00436-009-1468-z
Ye, J., McGinnis, S., Madden, T.L., 2006. BLAST: Improvements for better sequence analysis. Nucleic Acids Res. 34, 10–13. https://doi.org/10.1093/nar/gkl164
Yera, H., Estran, C., Delaunay, P., Gari-Toussaint, M., Dupouy-Camet, J., Marty, P., 2006. Putative Diphyllobothrium nihonkaiense acquired from a Pacific salmon (Oncorhynchus keta) eaten in France; genomic identification and case report. Parasitol. Int. 55, 45–49. https://doi.org/10.1016/j.parint.2005.09.004
Yera, H., Nicoulaud, J., Dupouy-Camet, J., 2008. Use of nuclear and mitochondrial DNA PCR and sequencing for molecular identification of *Diphyllobothrium* isolates potentially infective for humans. Parasite 15, 402–407. https://doi.org/10.1051/parasite/2008153402
Young, N.D., Nagarajan, N., Lin, S.J., Korhonen, P.K., Jex, A.R., Hall, R.S., Safavi-Hemami, H., Kaewkong, W., Bertrand, D., Gao, S., Seet, Q., Wongkham, S., Teh, B.T., Wongkham, C., Intapan, P.M., Maleewong, W., Yang, X., Hu, M., Wang, Z., Hofmann, A., Sternberg, P.W., Tan, P., Wang, J., Gasser, R.B., 2014. The Opisthorchis viverrini genome provides insights into life in the bile duct. Nat. Commun. 5. https://doi.org/10.1038/ncomms5378

Zambrano-Villa, S., Rosales-Borjas, D., Carrero, J.C., Ortiz-Ortiz, L., 2002. How protozoan parasites evade the immune response. Trends Parasitol. 18, 272–278. https://doi.org/10.1016/S1471-4922(02)02289-4

Zbikowska, E., Nowak, A., 2009. One hundred years of research on the natural infection of freshwater snails by trematode larvae in Europe. Parasitol. Res. 105, 301–311. https://doi.org/10.1007/s00436-009-1462-5

Zhang, L., Wu, Q., Hu, Y., Wu, H., Wei, F., 2015a. Major histocompatibility complex alleles associated with parasite susceptibility in wild giant pandas. Heredity 114, 85–93. https://doi.org/10.1038/hdy.2014.73

Zhang, L., Wu, Q., Hu, Y., Wu, H., Wei, F., 2015b. Major histocompatibility complex alleles associated with parasite susceptibility in wild giant pandas. Hered. Edinb 114, 85–93. https://doi.org/10.1038/hdy.2014.73

Ziętara, M.S., Kuusela, J., Veselov, A., Lumme, J., 2007. Molecular faunistics of accidental infections of Gyrodactylus Nordmann, 1832 (Monogenea) parasitic on salmon Salmo salar L. and brown trout Salmo trutta L. in NW Russia. Syst. Parasitol. 69, 123–135. https://doi.org/10.1007/s11230-007-9121-7

Zueva, K.J., Lumme, J., Veselov, A.E., Kent, M.P., Lien, S., Primmer, C.R., 2014. Footprints of directional selection in wild atlantic salmon populations: Evidence for parasite-driven evolution? PLoS ONE 9. https://doi.org/10.1371/journal.pone.0091672

10 Appendix

Codon#	Codon Start	Triplet	dS	dN	dN-dS	P-value	Normalized dN-dS	Translate	Amino Acid
18	52	TGT	0	0.4210573	0.42105728	0.888733477	0.799695252	C	Cys
19	55	TTT	0	0	0	N/A	0	F	Phe
20	58	CAA	0	0	0	N/A	0	G	Gly
21	64	GGA	4.060813	0.4988136	-3.56199923	0.996118156	-6.765145769	G	Gly
22	70	ATA	0	4.7349172	4.73491721	0.071173391	8.992816422	K	Lys
23	73	TAT	0	3.1267649	3.12676491	0.426391529	5.938524703	Y	Tyr
24	76	CAG	5.50952	0.4246729	-5.08484688	0.997617051	-9.657422176	Q	Gln
25	79	ATT	1.610151	0	-1.61015119	1	-3.058088114	I	Ile
26	82	CAA	0	0.4325112	0.43251122	0.850729146	0.821449205	Q	Gln
27	85	CAG	0	0	0	N/A	0	Q	Gln
28	88	CAC	0	0	0	N/A	0	H	His
29	91	CGT	0	1.8486262	1.8486262	0.768584505	1.611757782	R	Arg
30	94	GGG	0	0	0	N/A	0	G	Gly
31	97	GAA	0	0	0	N/A	0	E	Glu
32	100	GTA	6	3.6491001	3.6491001	0.426391529	5.938524703	V	Val
33	103	TGT	0	0	0	N/A	0	C	Cys
34	106	TGG	0	0	0	N/A	0	W	Trp
35	109	ATA	0	0	0	N/A	0	I	Ile
36	112	CAC	0	0	0	N/A	0	H	His
37	118	GCA	0	1.4919286	1.49192862	0.301131275	2.833553282	A	Ala
38	121	TGG	0	0.441071	0.441071	0.990789715	0.837706424	W	Trp
39	124	AGT	0	0.3692272	0.36922718	0.903308736	0.701256673	S	Ser
40	127	GAA	0	2.8032511	2.80325115	0.184671182	5.324089486	Q	Glu
41	130	GAA	0	1.7474866	1.74748664	0.518234689	3.318923197	Q	Glu
42	133	TGC	0	0	0	N/A	0	C	Cys
43	136	AGA	0	0.5	0.5	0.743810449	0.94962763	R	Arg
44	139	AGC	0	0	0	N/A	0	S	Ser
45	142	CTG	0	0.5842477	0.58424766	0.570534307	1.10963544	L	Leu
46	145	GAA	0	0	0	N/A	0	E	Glu
47	148	CCA	0	4.4358334	4.27062187	0.996184183	4.111001049	P	Pro
48	163	CCG	1.034886	2.9502736	1.91538787	0.28461396	3.637810494	P	Pro
49	166	GGC	0	0.5	0.5	0.666666667	0.94962763	G	Gly
50	172	AGA	0	4.4523385	4.45233852	0.068436725	8.456127354	R	Arg
51	175	GCG	3	1	-2	0.95473251	-3.798510521	A	Ala
52	178	AGG	0	3.0574329	3.39297512	0.775640653	3.746360055	R	Arg
53	184	GCT	1	0	-1	1	-1.89925526	A	Ala
54	187	GGA	0	0.5777357	0.57773566	0.645432724	1.097267486	G	Gly
55	190	GCG	0	0.502645	0.50264497	0.665493194	0.954651101	A	Ala
56	202	GCC	2	1.5008317	-0.49916828	0.789940911	-0.948047975	A	Ala
57	208	CGC	0	0	0	N/A	0	R	Arg
58	214	TAT	0	0	0	N/A	0	Y	Tyr
59	220	CTA	0	0	0	N/A	0	L	Leu
60	223	CAG	0	0	0	N/A	0	Q	Gln
61	226	CGC	0	0	0	N/A	0	R	Arg

Supplementary table 1 : table denoting selection data from translated Western Loch consensus protein sequence. Green; positive selection, Blue; Purifying selection and yellow; neutral

Codon#	Codon Start	Triplet	dS	dN	dN-dS	P-value	Normalized dN-dS	Translated	Amino Acid
20	58	CAA	0	0.4399527	0.43995266	0.838042967	1.65685887	Q	Glu
21	64	AAG	0	1.7375276	1.73752757	0.54247313	6.543517581	K	Lys
22	67	GCT	1	0	-1	1	-3.76599353	A	Ala
23	70	GAA	0	1.7678399	1.76783993	0.487644337	6.657673748	E	Glu
24	73	TAT	0	1.3262434	1.32624342	0.698201224	4.994624156	Y	Tyr
25	76	GTC	0	0.9800567	0.9800567	0.462716601	3.690887179	V	Val
26	79	AGA	0	0	0	N/A	0	R	Arg
27	82	TTC	0	0	0	N/A	0	F	Phe
28	85	AAC	0	0	0	N/A	0	N	Asn
29	88	AGC	0	0.3724097	0.37240974	0.895071474	1.402492665	S	Ser
30	91	ACT	0	0.5	0.5	0.666666667	1.882996765	T	Thr
31	94	GTG	0	0	0	N/A	0	V	Val
32	97	GGG	0	0	0	N/A	0	G	Gly
33	100	AAG	0	0	0	N/A	0	K	Lys
34	103	TTT	0	1.8807839	1.88078388	0.584715625	7.083019923	F	Phe
35	106	GTT	0	0	0	N/A	0	V	Val
36	109	GGA	0	0	0	N/A	0	G	Gly
37	115	CAC	0	0	0	N/A	0	H	His
38	121	GTA	1	1	0	0.740740741	0	V	Val
39	124	TGG	0	0	0	N/A	0	W	Trp
40	127	AGT	0	0	0	N/A	0	S	Ser
41	130	GTA	1.309797	0.4730978	-0.83669961	0.929587422	-3.151005335	V	Val
42	133	CAA	0	0.4347313	0.43473129	0.850852024	1.637195218	Q	Gln
43	136	TGC	0	0.4245183	0.42451831	0.888886609	1.598733196	C	Cys
44	139	AGA	0	0.5	0.5	0.746319784	1.882996765	R	Arg
45	142	AGC	0	0	0	N/A	0	S	Ser
46	148	GAA	0	0	0	N/A	0	E	Glu
47	151	CAA	0	0	0	N/A	0	Q	Gln
48	154	AGG	0	0.8567274	0.85672736	0.609264025	3.226429692	R	Arg
49	157	TCC	0	0	0	N/A	0	S	Ser
50	166	GCT	0	0	0	N/A	0	A	Ala
51	169	GGC	0	0	0	N/A	0	G	Gly
52	175	AGA	0	2.4666843	2.46668434	0.240661112	9.289517259	R	Arg
53	178	GCT	0	0	0	N/A	0	A	Ala
54	181	TGG	0	0.8918304	0.89183045	0.989755953	3.358627694	W	Trp
55	187	AGC	0	0	0	N/A	0	S	Ser
56	190	TGG	0	0.4459604	0.4459604	0.991782617	1.67948398	W	Trp
57	193	AGC	0	0	0	N/A	0	S	Ser
58	202	GTA	0	0	0	N/A	0	V	Val
59	205	AGC	0	0.3833748	0.38337477	0.869471232	1.443786903	S	Ser
60	208	ATA	1.403774	0.4371328	-0.96664081	0.94361503	-3.640363055	I	Ile
61	211	ACG	0	0.5	0.5	0.666666667	1.882996765	T	Thr
62	217	CTA	0	0.5731533	0.57315334	0.581577929	2.158491775	L	Leu
63	220	TTG	1.306459	1.5065629	0.20010383	0.690849506	0.753589732	L	Leu
64	223	ACT	0	0	0	N/A	0	T	Thr
65	226	ACA	0	0	0	N/A	0	T	Thr
66	229	GCG	0	0	0	N/A	0	A	Ala
67	232	CCA	0	2.546952	2.54695202	0.097964411	9.591804821	P	Pro

Supplementary table 2 : table denoting selection data from translated Elevated Loch consensus protein sequence. Green; positive selection, Blue; Purifying selection and yellow; neutral

Population	Tajima's D	Fu's F	Raggedness
AAG	-0.59475	-2.365	*r*= 0.01326
FES	-0.55609	-0.14829	
LAM	-1.25847	-0.716	
LAP	-1.08655	-8.045	

LAS	-0.67905	-0.1
LB	-1.96925	-1.506
LCA	-0.81643	-0.512
LDB	-1.2069	-0.508
LFD	-0.49027	-1.57
LFE	-0.18319	-0.19117
LFM	-0.61857	-0.409
LGD	-0.20664	-0.401
LMD	-0.14354	-6.882
LNB	-0.32024	-2.723
LNO	-0.83027	-1.906
LNU	-0.43673	-1.406
TWA	-0.61142	-0.477
Overall	-1.32785	-3.8852
P value	0.041	0.011

	Tajima's D	***Fu's F***	**Raggedness**
Elevated Loch system	-1.28251 $P = 0.90$	-20.108	0.0207
Western Loch	-0.8252 $P = 0.883$	-12.529	0.0286

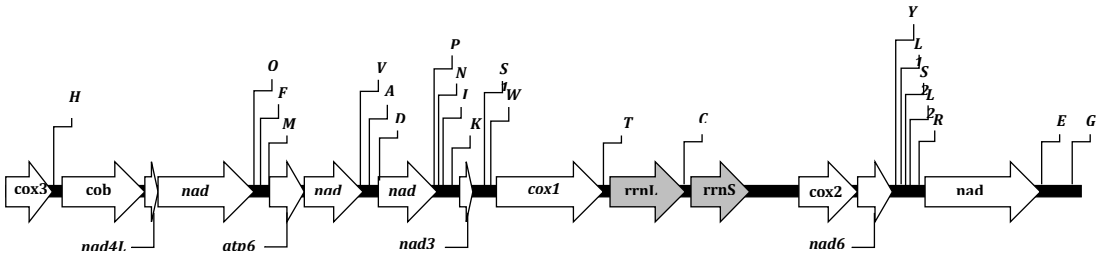
Supplementary table 3; Tajima's *D* and Fu's *F* calculation results for each population with overall figure, the overall output from both equations are tested using *P* values for significance of Tajima's *D* and Fu's *F* output. Further calculation output of Raggedness calculation: *r* value.

Name	Start	Stop	Length	codon
cox3	1	651	471	
trnaH	680	747	68	GTG
cob	751	1803	1053	
nad4L	1863	2117	255	
nad4	2192	3262	1071	
trnaQ	3386	3450	65	TTG
trnaF	3461	3525	65	GAA
trnaM	3541	3600	60	CAT
atp6	3696	4109	414	
nad2	4153	5016	864	
trnaV	5045	5107	63	TAC
trnaA	5120	5185	66	TGC
trnaD	5202	5266	65	GTC
nad1	5312	6130	819	
trnP	6177	6241	65	TGG
trnaN	6247	6312	66	GTT
trnaL	6319	6384	66	GAT
trnaK	6385	6450	66	CTT
nad3	6571	6801	231	
trnaS	6814	6874	61	GCT
trnaW	6884	6951	68	TCA
cox1	6960	8462	1503	
trnaT	8636	8703	68	TGT
rrnL	8710	9683	973	
trnaC	9696	9762	69	GCA
rrnS	9764	10495	734	
cox2	10525	11118	594	
nad6	11143	11565	423	
trnaY	11616	11680	65	GTA
trnaL	11688	11756	69	TAG
trnaS	11757	11823	67	TGA
trnaL	11841	11905	65	TAA
trnaR	11946	12013	68	TCG
nad5	12044	13156	1113	
trnE	13338	13401	62	
trnG	13598	13668	70	

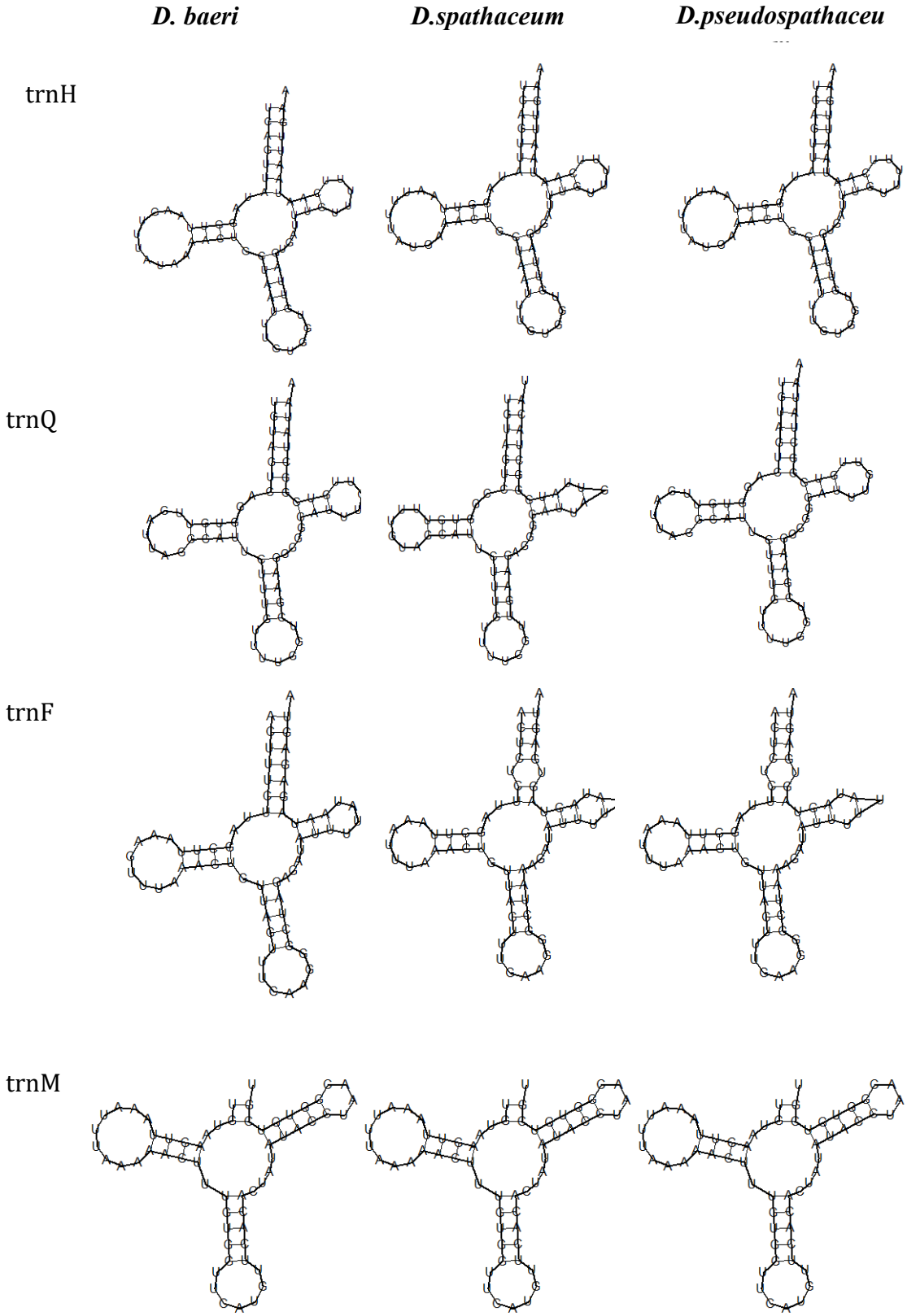
Supplementary Table 4: Summary table of the mitochondrial protein coding genes in *D. baeri* including position start, stop, and length of predicted gene and terminating codon.

Gene	Length, bp			Position,5' to 3'		
	D.baeri	D.spathaceum	D.pseudospathaceum	D.baeri	D.spathaceum	D.pseudospathaceum
cox3	651	655	655	1-651	1-655	1-655
trnaH	68	68	68	680-747	680-747	680-747
cob	1053	1111	1111	751-1803	751-1861	751-1861
nad4L	255	264	264	1863-2117	1863-2126	1863-2126
nad4	1071	1299	1296	2192-3262	2087-3385	2087-3382
trnaQ	65	64	64	3386-3450	3389-3452	3385-3448
trnaF	65	65	64	3461-3525	3465-3529	3458-3521
trnaM	60	69	69	3541-3600	3566-3634	3532-3600
atp6	414	519	519	3696-4109	3638-4156	3604-4122
nad2	864	892	901	4153-5016	4182-5073	4139-5039
trnaV	63	63	63	5045-5107	5074-5136	5040-5102
trnaA	66	67	64	5120-5185	5146-5212	5115-5178
trnaD	65	65	65	5202-5266	5246-5130	5212-5276
nad1	819	910	910	5312-6130	5312-6221	5277-6186
trnP	65	65	65	6177-6241	6222-6286	6187-6251
trnaN	66	67	66	6247-6312	6290-6356	6255-6320
trnaL	66	66	66	6319-6384	6363-6428	6325-6390
trnaK	66	69	68	6385-6450	6430-6498	6393-6460
nad3	231	357	357	6571-6801	6502-6858	6465-6821
trnaS	61	60	60	6814-6874	6862-6921	6824-6883
trnaW	68	67	68	6884-6951	6931-6997	6893-6960
cox1	1503	1659	1659	6960-8462	7007-8665	6970-8628
trnaT	68	65	65	8636-8703	8694-8758	8659-8723
rrnL	734	992	1000	9323-9683	8759-9750	8724-9723
trnaC	69	70	71	9696-9762	9751-9820	9724-9794
rrnS	361	725	725	9764-10495	9821-10545	9795-10519
cox2	594	615	615	10525-11118	11355-11969	10548-11162
nad6	423	459	459	11143-11565	11984-12442	11175-11633
trnaY	65	64	64	11616-11680	12457-12520	11648-11711
trnaL	69	70	68	11688-11756	12525-12594	11720-11787
trnaS	67	67	67	11757-11823	12595-12661	11788-11854
trnaL	65	67	67	11841-11905	12681-12747	11877-11943
trnaR	68	67	68	11946-12013	12784-12850	11979-12046
nad5	1113	1587	1587	12044-13156	12851-14437	12047-13633
trnE	62	62	66	13338-13401	14459-14520	13653-13718
trnG	70	71	64	13598-13668	14709-14779	14031-14093

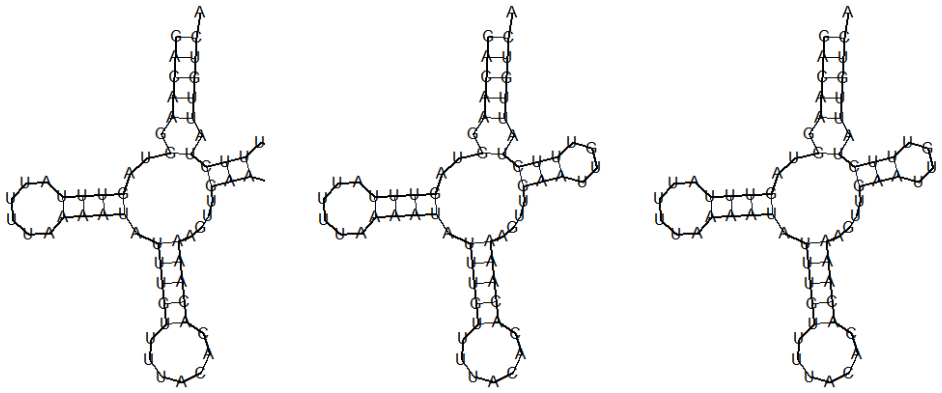
Supplementary Table 5: Summary data of mt genome organization of the *D. baeri* mt genome in comparison with previously published *D.spathaceum* and *D.pseudospathaceum,* comparisons include; position, sequence length and terminating codons across all annotated genes in all three mt genomes



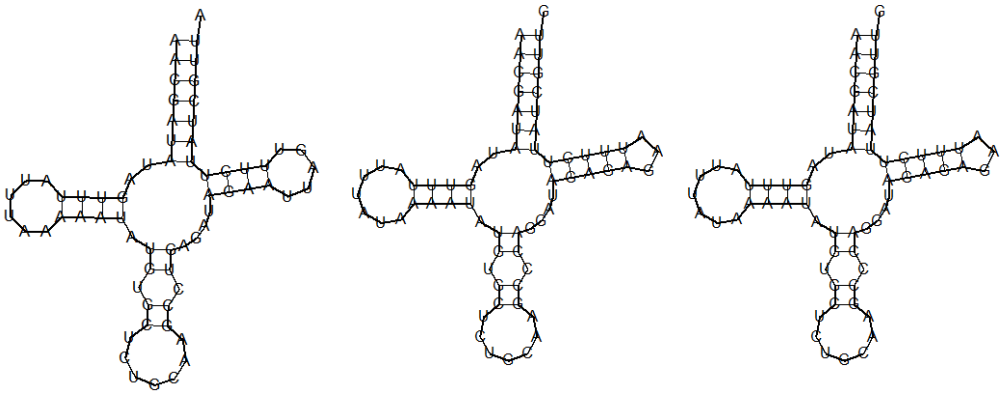
Supplementary figure 1: Linearized construction of *D. baeri* mitochondrial genome with protein coding genes, trna and rrna order.


D. baeri
D.spathaceum
D.pseudospathaceu
trnH
trnQ
trnF
trnM

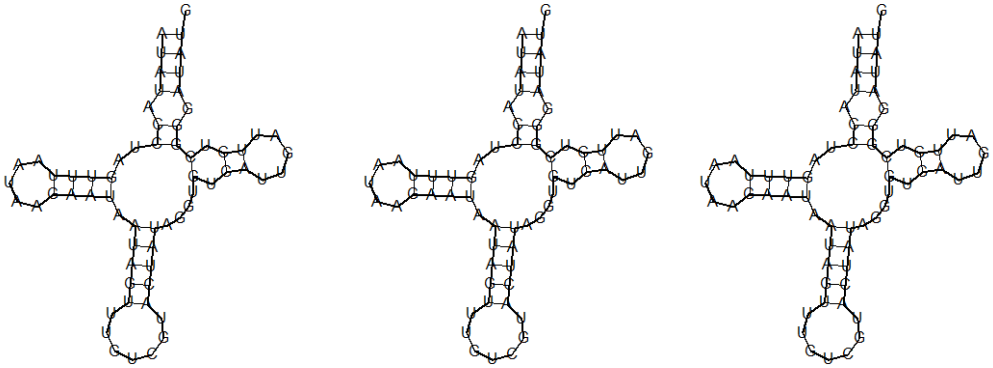
trnV



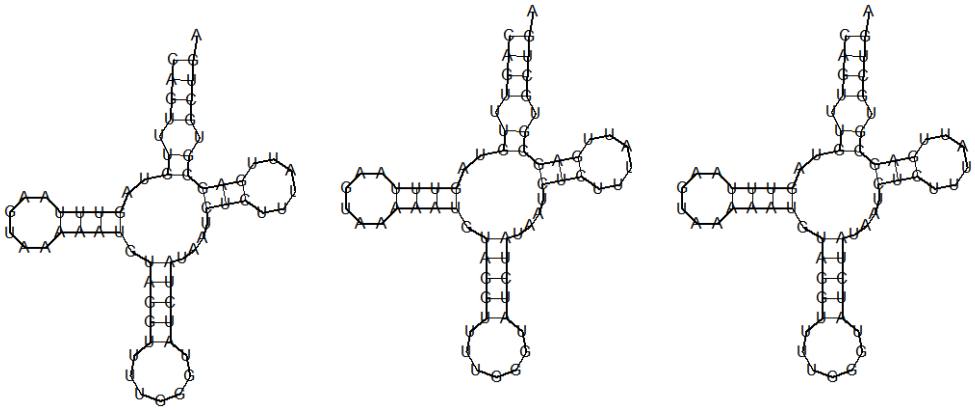
trnA



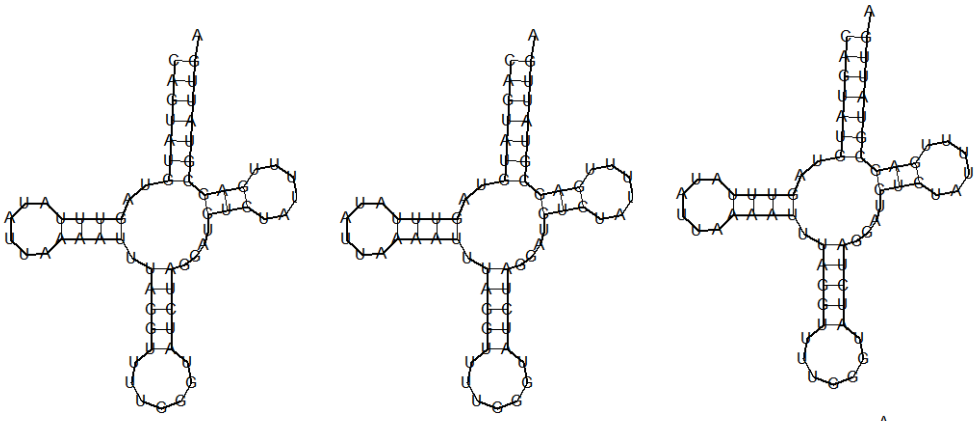
trnD



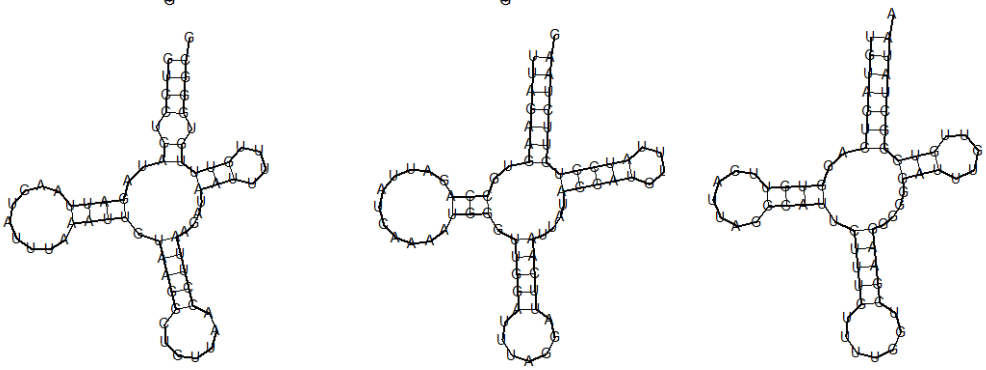
trnP



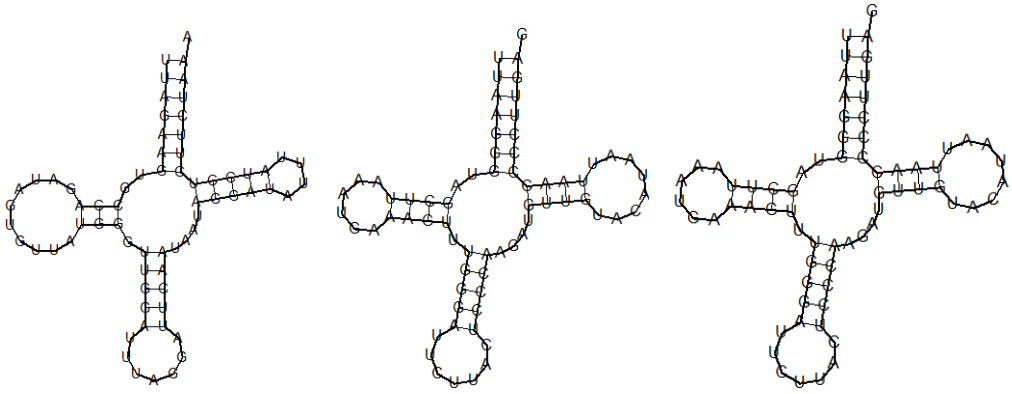
trnN



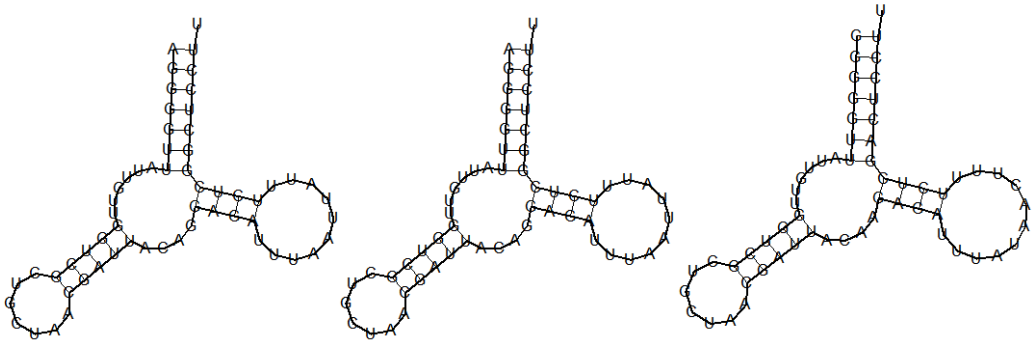
trnL1



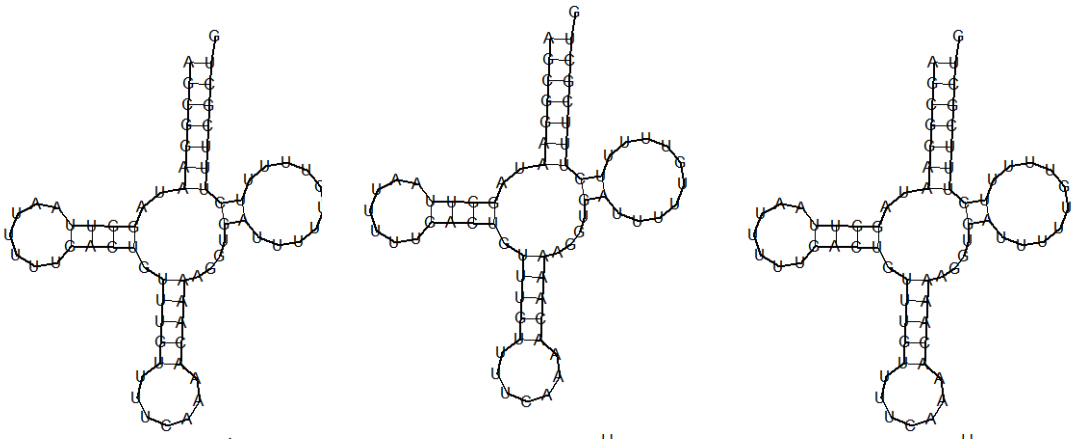
trnK



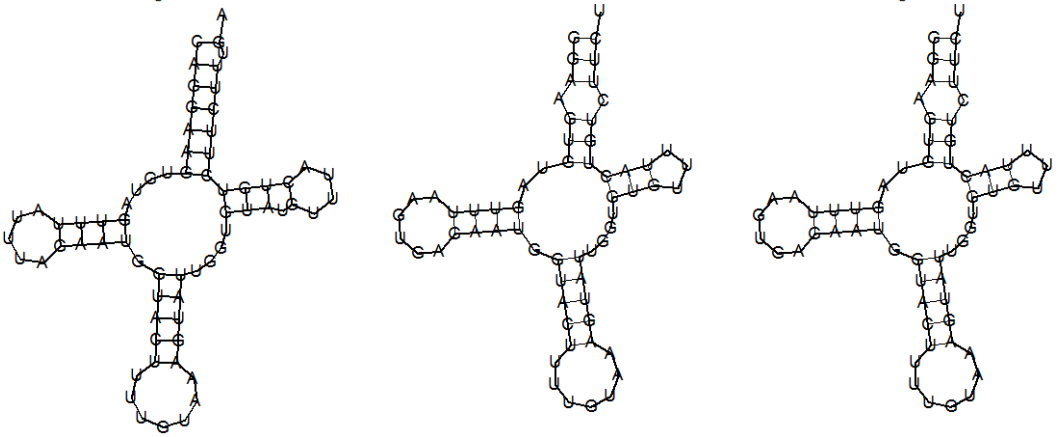
trnN



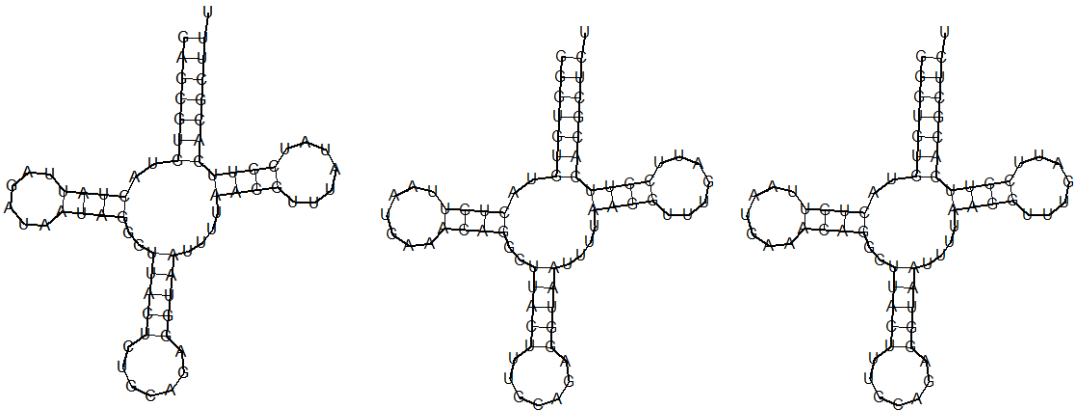
trnS



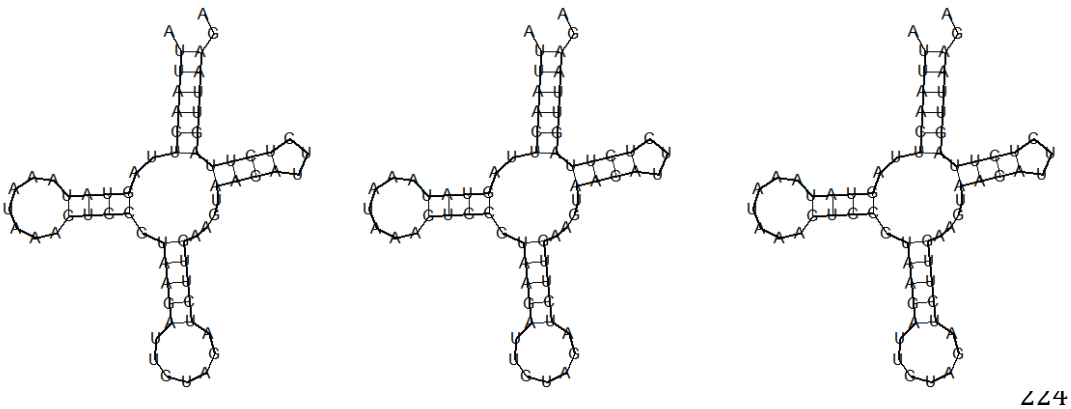
trnW



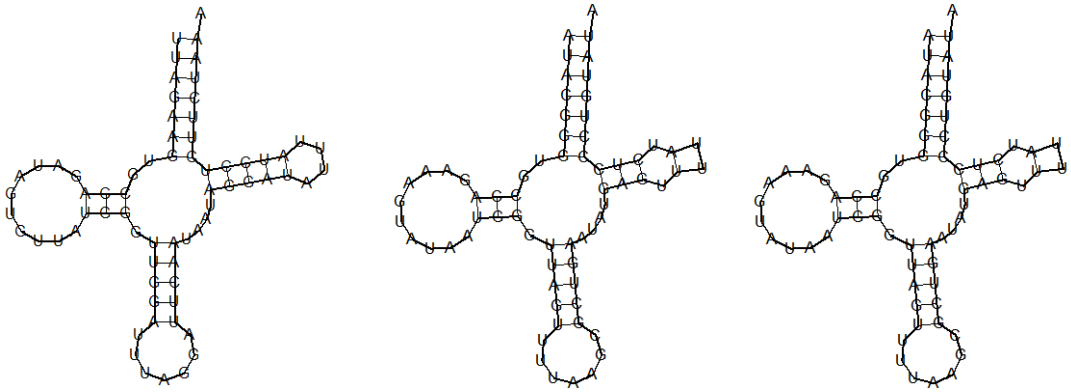
trnT



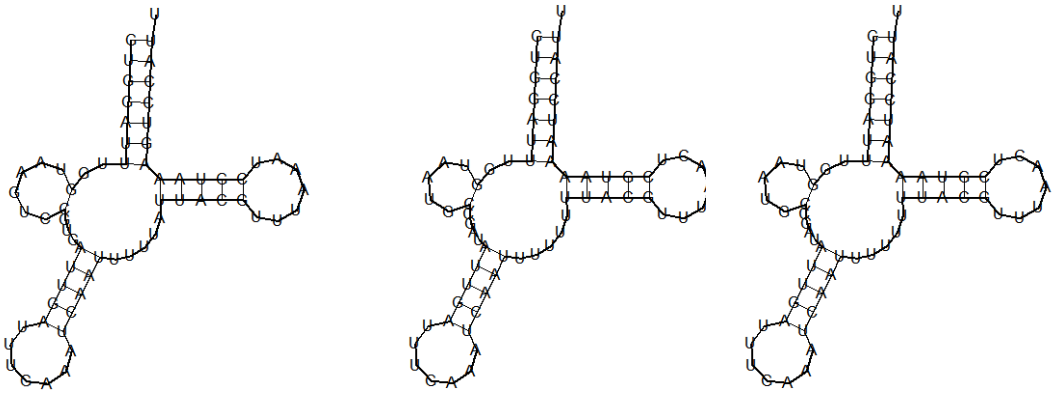
trnC



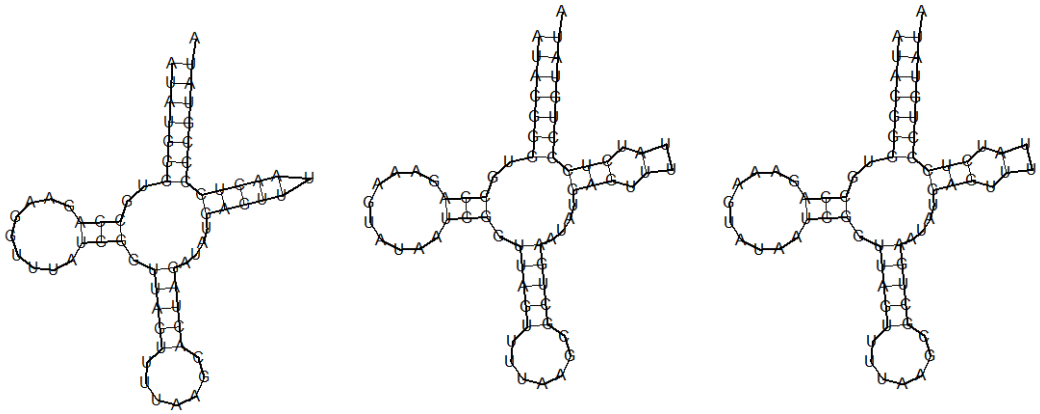
trnY



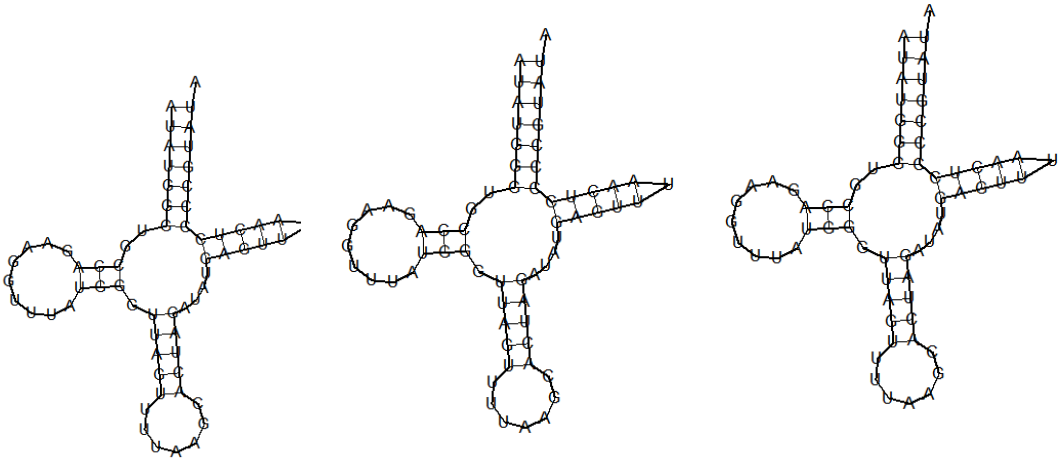
trnS2

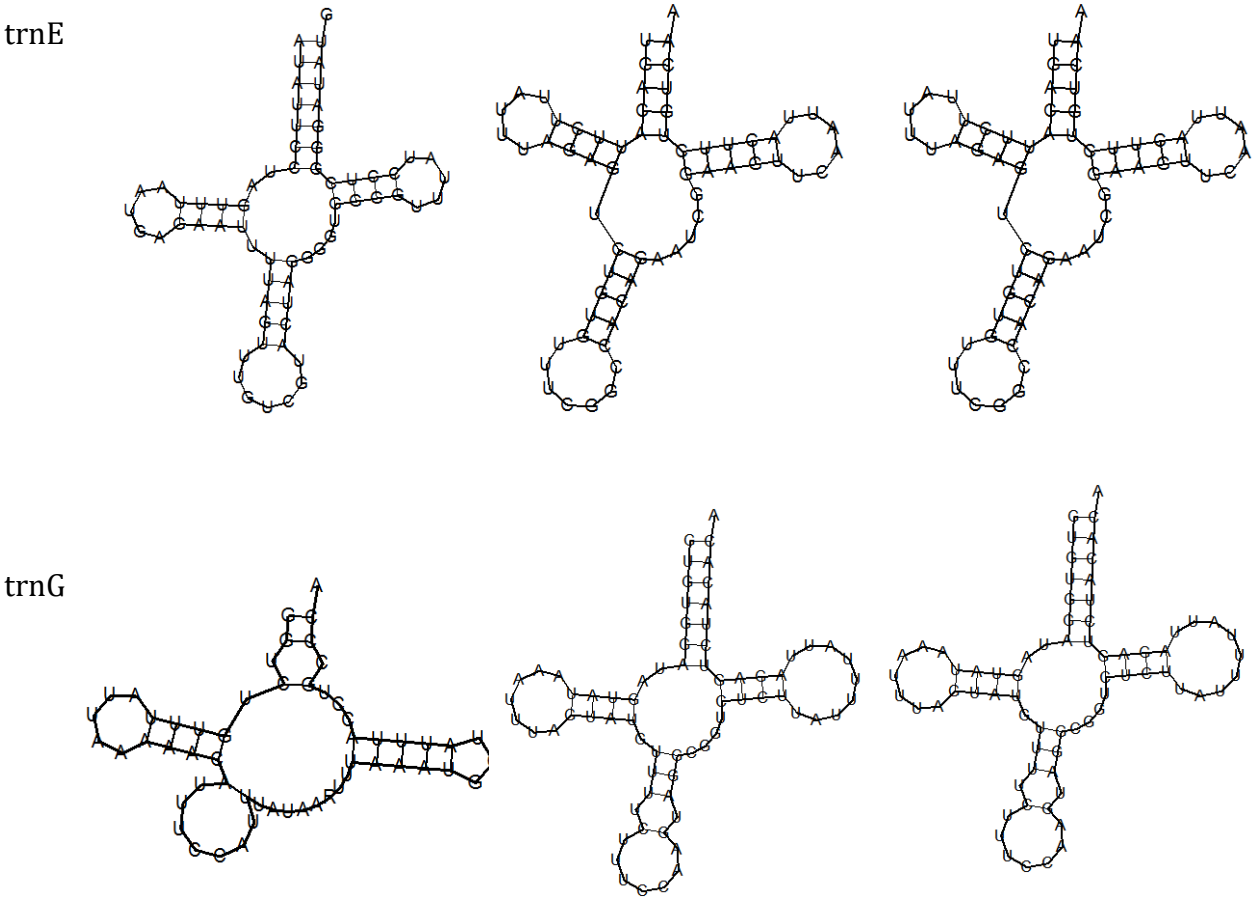


trnL2



trnR





Supplementary figure 2: tRNA modeling within *Diplostomum* genomes columns denoted with species; left *D. baeri,* centre *D.spathaceum*, right *D.pseudospathaceum*. Construction of RNA was performed using RNA fold server